

# Exploring the Role of Phosphorylation in *Arabidopsis thaliana* Pollination-Specific Tissue by Applying Phosphoproteomics

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

**Pururawa Mayank**

aus

Indien

**Promotionskomitee**

Prof. Dr. Ueli Grossniklaus

Prof. Dr. Christian von Mering

Dr. Thomas Nuhse

**Zürich, 2013**

**Dedicated to my Parents**

**Madan Chaubey**

**And**

**Chandra Prabha Chaubey**

## ***Zusammenfassung***

---

Erfolgreiche Bestäubung ist abhängig von der Zell-Zell-Kommunikation und einer schnellen, zellulären Antwort. *Arabidopsis thaliana* gehört zu der Gruppe der Selbstbestäuber. Die Bestäubung beginnt mit der Ablage der dehydrierten, ausgereiften Pollenkörner auf dem Stigma. Nach der Rehydrierung beginnt die Keimphase und es entsteht ein Pollenschlauch, durch den die zwei Spermien zum weiblichen Gametophyten (Embryosack) transportiert werden, so dass Doppelbefruchtung stattfinden kann. Verschiedene Studien zu diesem Thema haben belegt, dass ausgereifte und dehydrierte Pollenkörner alle für die Keimung und Initiierung des Pollenschlauchwachstums nötigen Transkripte und Proteine bereits enthalten. Aus diesem Grund ist es wichtig, die Rolle der posttranslationalen Modifikationen (hier Phosphorylierung), die möglicherweise viele durch Pollinierung induzierte Prozesse kontrolliert, zu untersuchen.

Phosphorylierung ist einer der Schlüsselmechanismen der posttranslationalen Modifikation zur Regulation von Proteinfunktionen in diversen biologischen Mechanismen und Kontexten. Sie kann die Proteinaktivität verändern, indem sie Proteine, die einen einzelnen Mechanismus kontrollieren, aktiviert oder deaktiviert. Dafür ist es nicht ausreichend zu wissen, ob ein Protein phosphoryliert ist oder nicht. Es ist vielmehr wichtiger die exakte Aminosäure, die unter spezifischen Bedingungen phosphoryliert wird, zu kennen. Im Allgemeinen enthält diese Dissertation Informationen über die mögliche Rolle und das Auftreten von phosphorylierten Proteinen, während der Bestäubung von *Arabidopsis thaliana*.

In Kapitel 3 der Arbeit wird eine phosphoproteomische Analyse von vollentwickelten *Arabidopsis*-Pollenkörnern durchgeführt. Um potentielle Ziele der Phosphorylierung zu identifizieren wurden drei verschiedene Anreicherungschromatographien durchgeführt und dabei eine weitreichende Abdeckung der Pollen-Phosphoproteine mit 962 phosphorylierten Peptiden, die zu 598 Phosphoproteinen korrespondieren, erhalten. Zusätzlich konnten 609 bereits nachgewiesene Phosphorylierungsstellen bestätigt werden. Ein Abgleich mit der PhosPhat Datenbank, die empirische Daten zum *Arabidopsis*-Phosphoproteom enthält, ergab dass 207 der 240 nicht in der Datenbank erfassten Phosphoproteine weisen eine hohe Expression in Pollen aufweisen. Daraufhin zeigte sich bei einer genontologischen (GO) Anreicherungsanalyse, dass für diese 240 Phosphoproteine eine Überrepräsentation, der für das Pollenschlauchwachstum kritischen GO Kategorien vorliegt. Dies deutet darauf hin, dass Phosphorylierung spätere Mechanismen der Pollenentwicklung reguliert. Mit Hilfe von Phosphopeptid-Motivanalysen konnte eine Überrepräsentation von  $\text{Ca}^{2+}$ /Calmodulin abhängigen bzw. Mitogen-aktivierten

Proteinkinasmotiven und Bindungsmotiven von 14-3-3 Proteinen aufgezeigt werden. Schliesslich konnte eine Tyrosin-Phosphorylierungsstelle identifiziert werden, die eine Bestätigung des TDY dualen Phosphorylierungsmotivs der Mitogen-aktivierten Proteinkinase darstellt (MPK8/MPK15). Diese Studie liefert somit eine solide Basis, für weitere Untersuchungen der Rolle von Phosphorylierungen, während der Pollenentwicklung.

In Kapitel 4 wird die qualitative und quantitative Analyse von Phosphoproteinen/-peptiden im Zusammenhang mit der Pollinierung beschrieben. Hierfür wurden pollinierungsspezifische Gewebe, wie z.B. Pollenkörner und unbestäubtes weiblichen Gewebe, bzw. Weithiches Gewebe eine Stunde nach der Bestäubung, verglichen. Die Daten zeigen jeweils eine unterschiedliche Expression von einigen Phosphopeptiden zwischen Pollen und Bestäubten weiblichem Gewebe, wie auch zwischen unpolliniertem und eine Stunde nach der Pollinierung gesammeltem weiblichen Gewebe.

Einige Phosphopeptide ändern ihren Phosphorylierungsstatus zwischen Pollen und polliniertem weiblichem Gewebe, was neue Einsichten in die während der selbtskompatiblen Befruchtung stattfindende Initialkommunikation gewährt. Darüber hinaus liefern diese daten die grandlage für weitergehende untersuchungen, z.b. mit Hilfe der ortsspezifischen Mutagenese von Phosphorylierungsstellen, die spezifisch für die Pollinierung sind, die Relevanz der Phosphorylierung während der bestäubung aufzuklären.

## Summary

---

Successful pollination depends on cell-cell communication and rapid cellular responses. The flowers of the plant model *Arabidopsis thaliana* self-pollinate, and therefore the pollination event initiates when a dehydrated, mature pollen grain lands on the stigmatic surface of female tissue (pistil). Upon its arrival, the pollen grain forms a tight contact junction with the female tissue, which leads to pollen hydration, germination, and the formation of a pollen tube that elongates to carry two sperm cells to the female gametophyte (embryo sac) for double fertilization to ensue. Various studies have emphasized that a mature, dehydrated pollen grain contains all the transcripts and proteins required for germination and initial pollen tube growth. Therefore, it is important to explore the role of post-translational modifications (here phosphorylation), through which probably many processes induced by pollination are regulated.

Phosphorylation is one of the key mechanisms of post-translational modifications that regulate protein function in diverse biological pathways and contexts. Phosphorylation can alter the protein activity and hence activates or deactivates proteins controlling a specific pathway. Therefore, it is not sufficient to just know whether a protein is phosphorylated or not, rather it is important to know which particular amino-acid residue of the protein is getting phosphorylated under given specific conditions. In general, this thesis contains information about the occurrence and possible role of phosphorylated proteins in the pollination process of *Arabidopsis thaliana*.

In Chapter 3 of this thesis, we report a phosphoproteomic study conducted on mature *Arabidopsis* pollen grains with the aim of identifying potential targets of phosphorylation. Using three enrichment chromatographies, a broad coverage of pollen phosphoproteins with 962 phosphorylated peptides corresponding to 598 phosphoproteins was obtained. Additionally, 609 confirmed phosphorylation sites were successfully mapped. 207 of 240 phosphoproteins that were absent from the PhosPhAt database which contains the empirical *Arabidopsis* phosphoproteome, showed highly enriched expression in pollen. Gene ontology (GO) enrichment

analysis of these 240 phosphoproteins shows an over-representation of GO categories crucial for pollen tube growth, suggesting that phosphorylation regulates later processes of pollen development. Moreover, motif analyses of pollen phosphopeptides showed an over-representation of motifs specific for  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinases, mitogen-activated protein kinases, and binding motifs for 14-3-3 proteins. Lastly, one unique tyrosine phosphorylation site was identified, validating the TDY dual phosphorylation motif of mitogen-activated protein kinases (MPK8/MPK15). This study provides a solid basis to further explore the role of phosphorylation during pollen development.

In Chapter 4, we discuss the qualitative and quantitative analysis of phosphoproteins/phosphopeptides associated with pollination. We compared the phosphoproteome of pollination-specific tissues, *i.e.* of unpollinated and pollinated male and female tissues. The data shows differential expression of the phosphopeptides between the pollen grain and female tissue unpollinated or 1HAP (one hour after pollination).

Overall, the studies presented in this thesis will lay the foundation for further analyses e.g. site-directed mutagenesis experiments, on pollination-specific phosphoproteins that may play key roles during plant reproduction.

# Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	<i>Arabidopsis thaliana</i> .....	1
1.1.1	The model organism of plant biology .....	1
1.2	Sexual reproduction and the life cycle .....	2
1.2.1	The Arabidopsis Male Gametophyte .....	4
1.2.2	The Arabidopsis Female Gametophyte.....	8
1.3	Pollination .....	10
1.3.1	General observations .....	10
1.3.2	Pollen-stigma interaction in compatible pollinations.....	11
1.3.3	Pollen-stigma interactions in self-incompatible species .....	12
1.4	Post Translational Modifications: Phosphorylation .....	13
1.4.1	History, discovery and importance .....	13
1.4.2	Functions, type of phosphorylation and identification .....	14
1.4.3	Role of Phosphorylation in pollination .....	15
1.4.4	Phosphoproteomics: Detection and Characterization .....	16
1.4.5	Detection of enriched phosphopeptides: Mass spectrometry.....	20
1.4.6	Quantification of Identified Phosphopeptides: .....	21
1.5	References .....	24
<b>2</b>	<b>Aim of the Thesis .....</b>	<b>37</b>
<b>3</b>	<b>Materials and Methods .....</b>	<b>37</b>
3.1	NOTE:.....	37
3.2	Pollen collection .....	37
<b>4</b>	<b>Characterization of the phosphoproteome of mature <i>Arabidopsis</i> pollen.....</b>	<b>37</b>
4.1	Supporting Information .....	51
4.1.1	Supplemental Experimental Procedures .....	51
4.2	References .....	55
4.3	Supporting Figure and Table Legends.....	57
4.3.1	Supporting Table S1: Overview of Pollen Phosphoproteome .....	57
4.3.2	Supporting Table S2: List of Phosphoproteins Required for Normal Pollen Development.....	57
4.3.3	Supporting Table S3: List of Published Arabidopsis Pollen Mutants (15.09.2011). ....	57
4.3.4	Supporting Table S4: Pollen-enriched phosphoproteins.....	57
4.3.5	Supporting Table S5: List of Kinases Found in the Pollen Phosphoproteome.....	58
4.3.6	Supporting Table S6: List of RLKs Found in the Pollen Phosphoproteome. ....	58
4.3.7	Supporting Figure S1: Outline of Experimental Procedures.....	58

4.3.8	Supporting Figure S2: Representation of phosphopeptide enrichment in six different SCX fractions coupled with three different phospho-enrichment chromatographies.	60
5	<b>Phosphoproteome map of pollination specific tissue involved in pollen–stigma interaction in <i>Arabidopsis thaliana</i></b>	61
5.1	Summary	62
5.2	Introduction	63
5.3	Results and Discussion	65
5.3.1	An extended atlas of the pollen phosphoproteome identifies 414 novel phosphoproteins	65
5.3.2	Qualitative phosphoproteome of pollination-specific tissue	67
5.3.3	Over-represented GO categories from dde2-2 gynoecea 1HAP, have obvious role in pollen-stigma interaction	69
5.3.4	Pollination-specific cross talk: A label-free approach indicates dephosphorylation dominates phosphorylation	73
5.3.5	Dual phosphorylation of MPK6 at TEY motif	76
5.4	Experimental Procedures	78
5.4.1	Plant material and sample collection	78
5.4.2	Protein extraction	79
5.4.3	Phosphopeptide enrichment	79
5.4.4	Data acquisition, peptide identification and label-free quantification	80
5.5	References	81
5.6	Supporting Information	87
5.6.1	Supplement Figure S1 (a-d): Level plots of quantified candidates:	87
5.6.2	Supplement Tables:	87
6	<b>Conclusion and Outlook</b>	89
7	<b>Acknowledgments</b>	91
8	<b>Appendix</b>	93
8.1	Supplement Figure S1 (a-d): Level plots of quantified candidates:	93
8.2	Supplement Tables: Presented in the CD along with the thesis.	93



## 1.1 *Arabidopsis thaliana*

### 1.1.1 The model organism of plant biology

Despite the important contribution of plants in our everyday life, we still lack valuable information about them. Although there has been a revolution in the biological sciences in the past twenty years, there is still a great deal that remains to be discovered in plants.

In order to better understand them, we need to have more insights into their genetic information, which can be accomplished by using a model plant species. During the last three decades, *Arabidopsis thaliana* has become universally recognized as an important model system in the plant research community. It is a small flowering plant of the mustard family (Brassicaceae), native to Europe, Asia, and north-western Africa. Even though it is a wild, non-commercial member of the mustard family, it possesses several properties, which makes it an ideal model system for plant research. The small size (approximately 25 cm tall) and short life cycle (6-8 weeks) of *Arabidopsis* favours research in limited space and time

(Figure 1.1). *Arabidopsis* naturally self-pollinates but can be cross-pollinated and bears approximately 4-5000 seeds per plant

(Leutwiler *et al.*, 1984). The large progeny size is useful, as it enables reliable statistical analysis of segregation ratios during extensive genetic experiments. *Arabidopsis* has a relatively small genome size, which has been reported to vary between 110 to 135 megabase pair (Mbp) in different studies (Leutwiler *et al.*, 1984; Bennett and Smith, 1976; TAIR-Genome, 2000). The genome consists of five chromosome pairs, each containing about 120Mbp of DNA, excluding telomeric, nucleolar, and centromeric tandem repeats (TAIR-Genome, 2000), with relatively small amounts of repeat sequences forming approximately 10% of the total genome (Dean and Schmidt, 1995). This further confirms *Arabidopsis* potential as a model organism for molecular genetics studies and as a fundamental organism for understanding key aspects of the molecular biology in higher organisms.

In the year 1990, a multinational *Arabidopsis* genome project was established and with the hardship of scientists located in Europe, Asia, and the United States of America, the whole genome of *Arabidopsis* was sequenced in year 2000, describing a total of 25498 nuclear, 80 chloroplast and 60 mitochondrial genes encoding proteins. The completion of the project represents a milestone

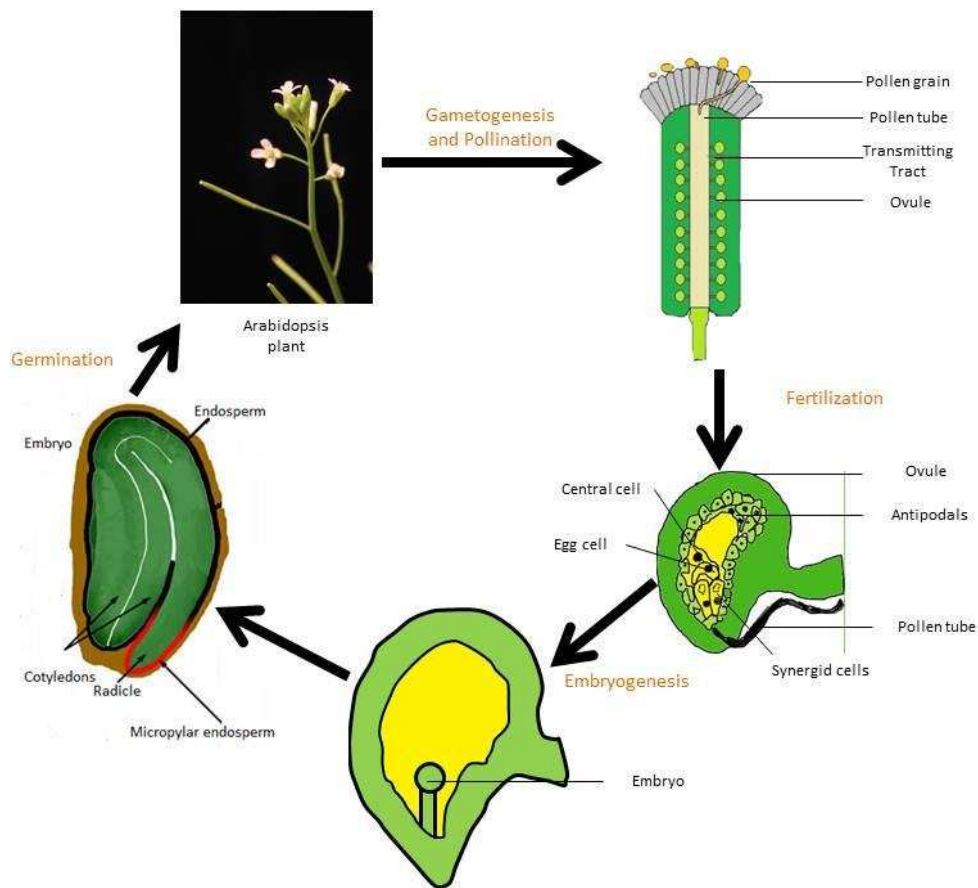


**Figure 1.1:** *Arabidopsis thaliana*.  
A wild type Col-0 plant at the  
flowering stage

in our understanding of the form and functions of not just this particular organism, but most of other plants, and has led to an improvement in several valuable crops that share important genetic information with *Arabidopsis*. Furthermore, the sequenced genome facilitated numerous genomics, transcriptomics, and proteomics studies on this model organism, which resulted in an increase of plant-specific journals.

## **1.2 Sexual reproduction and the life cycle**

The ability to reproduce is one of the defining criteria for life and can be broadly achieved both by asexual and sexual reproduction. Asexual reproduction occurs through mitotic division and, therefore, all the progeny are genetically identical to their parents. However, sexual reproduction, involving meiosis and syngamy (cell fusion), mixes the genetic information of the parents, resulting in progeny having a unique combination of parental traits. Thus, sexual reproduction yields new genetic combination which may be better adapted to a changing environment and, hence sexuality is the predominant mode of reproduction in animals and plants (Otto, 2003). The processes of meiosis and syngamy separate the life cycle into haploid (gametophyte) and diploid (sporophyte) phases (Coelho *et al.*, 2007). This is called a Hofmeister-Strasburger alternation of generations (Bell, 1994). In angiosperms, the sporophyte is the dominant generation, resulting in the formation of the haploid gametophyte, which in turn forms haploid gametes post meiosis (division that reduces the chromosome number by half). The haploid gamete fuses again to form a diploid zygote, which leads to the next sporophytic generation. The size and the life span of the diploid and haploid generations vary greatly amongst land plants.



**Figure 1.2:** Life cycle of an angiosperm (*Arabidopsis*): Alternation of sporophytic and gametophytic generations. Diploid sporophytic cells undergo meiosis, and further mitotic divisions that results in the formation of haploid gametophytic generation. The male gametophyte is called pollen, whose role is to carry two sperm cells to the female gametophyte encased in the pistil. The female gametophyte embryo sac is composed of seven cells. Fusion of one sperm cell with egg cell results in a diploid zygote and the other sperm cell fuses with the diploid central cell resulting in a triploid endosperm in a process known as double fertilization. (Adapted from Escobar-Restrepo *et al.*, 2007).

In angiosperms, the gametophyte, which develops within the sporophytic tissue, is extremely reduced in both life-span and size. Typically, a mature male gametophyte (pollen) contains three cells and is enclosed within the anther (sporophytic tissue) while the female gametophyte (FG) (embryo sac) contains seven cells, which are encased within the ovules enclosed inside the carpel (sporophytic tissue) (Grossniklaus and Schneitz, 1998). In the course of evolution, the FG became enclosed inside the sporophytic tissue, in order to protect the developing seed from drought and herbivorous insects. This inaccessibility of the FG required a new mode of sperm delivery which was achieved by the evolution of a pollen tube (PT) that delivers the sperm cells. The PT carries two sperm cells and penetrates the female sporophytic tissue (stigma) in order to bring the two sperm cells to the FG, where double fertilization occurs (Lopes and Larkins, 1993; Grossniklaus and Schneitz, 1998). One of the sperm cells fuses with the egg cell to form a diploid zygote, which later gives rise to an embryo. The second sperm cell fuses with the central cell of the FG and leads to the formation of the endosperm, a nutritive tissue that feeds the developing embryo or germinating seeds (Figure 1.2).

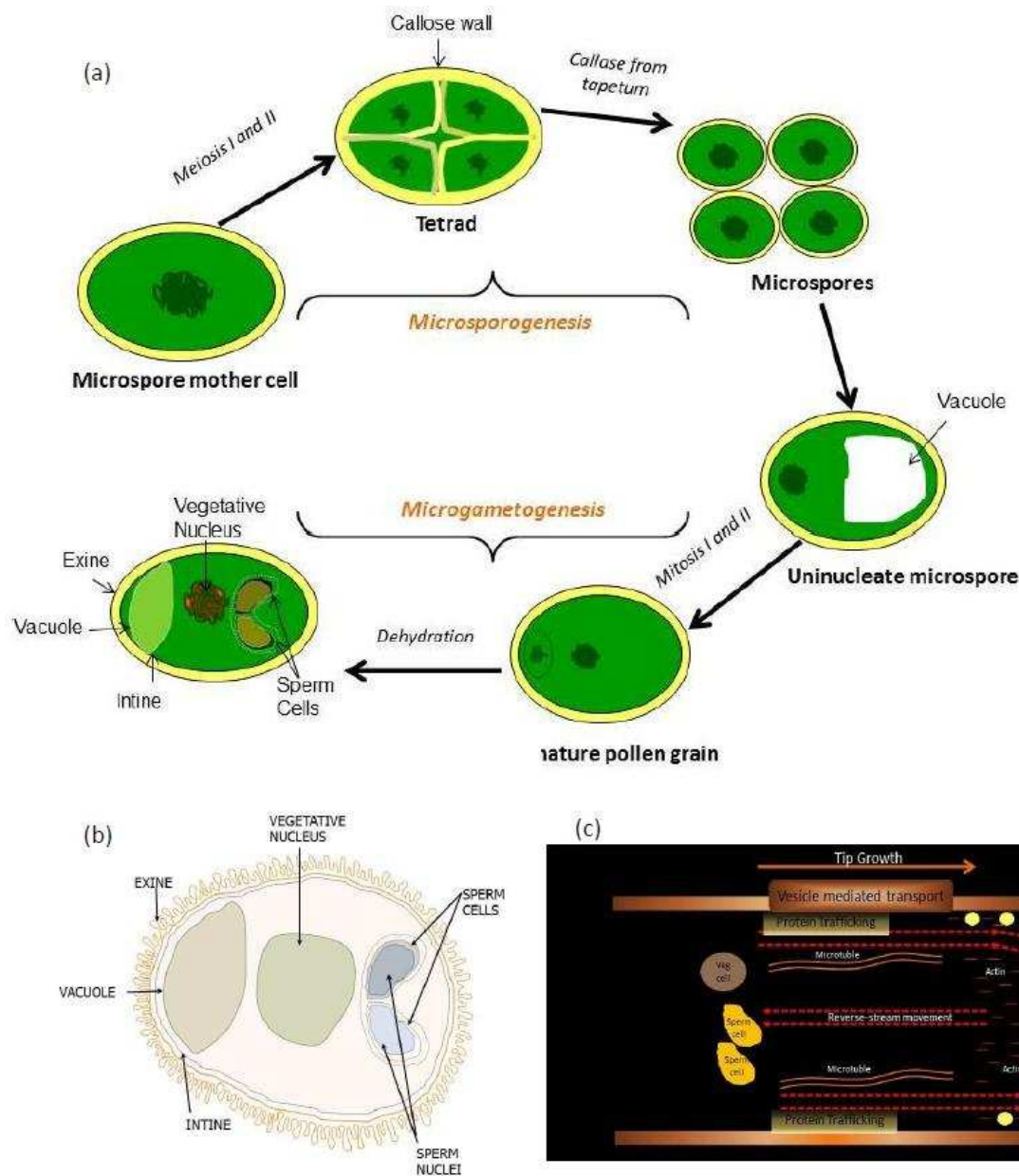
Existing land plants comprise the bryophytes (mosses, hornworts, and liverworts), tracheophytes (vascular or higher plants: club mosses, *horse tails*, ferns), gymnosperms (including the conifers) and angiosperms (flowering plants). Microfossil records show that all of these plant groups have evolved from a common bryophyte-like ancestor approximately 500 million years ago, in which gametophyte and sporophyte were both equal in life span and size. However, later in the course of evolution the sporophyte became more predominant while the gametophyte decreased in both life span and size (Graham and Wilcox, 2000; Kenrick, 2000).

The gametophyte of the basic vascular plants, e.g. a typical fern, is a small green microscopic structure, known as a prothallium, which is not dependent on the sporophyte for its nutrition and develops as an independent organism. The prothallium is often a hermaphroditic structure that produces both sperm and egg cells, fusion of which leads to the development of the zygote that later develops into a new sporophyte. The sporophyte produces small organs on the ventral side of the leaves, called sporangia, inside which cells undergo meiosis to produce haploid spores. After dispersal, the spores develop into a mature gametophyte. Fossil records suggest that during evolution the gametophyte was reduced and retained within the sporophyte (Ravan *et al.*, 1999). These plant “pro-gymnosperms” are known to be the progenitors of the existing seed plants, which can broadly be classified into gymnosperms and angiosperms.

## 1.2.1 The Arabidopsis Male Gametophyte

### 1.2.1.1 Development and structure of the pollen grain

Valerius Cordus coined the word “pollen” in the sixteenth century, after he noticed the dust particles omitted from flowers. In all angiosperms, the pollen is formed within the pollen sac of the anther and represents the male gametophyte. The cells of the sporogenic tissue within the pollen sacs undergo meiosis, forming four haploid microspores, which are encased within the cell wall of the microsporocyte mother cell. The sporophytic nutritive cell layer that surrounds the microspores (tapetum) releases an enzyme that digests the callose wall to release the haploid microspores. The formation of microspores is known as microsporogenesis (McCormick, 2004). The first asymmetric mitotic division of the haploid microspores generates a large vegetative cell and a small generative cell (Figure 1.3 a). The second mitotic division of the generative cell results in the formation of two sperm cells (male gametes). The second mitotic division is species-dependent, in case of *Arabidopsis* and maize it takes place before the pollen is released from the anther; however, in lily and tomato, the generative cells divide only after the pollen tube (PT) is formed.



**Figure 1.3 (a):** Illustrates the successive events leading to the formation of a mature pollen grain. The microspore mother cell (MMC) undergoes meiosis, which results in a tetrad of haploid microspores. The tetrads are then released as free microspores by the action of callase produced by the tapetum layer of the anther. These uninucleate microspores undergo an asymmetric mitotic division, resulting in a pollen grain containing two cells: a large vegetative cell and a small generative cell that is enclosed within the vegetative cell. The second mitotic division of generative cells results in two sperm cells that contribute the parental genome to the progeny. **(b)** Enlarged schematic of a mature dehydrated *Arabidopsis* pollen grain, adapted from: Mayank.P *et al* 2012. **(c)** Schematic of *Arabidopsis* pollen tube: Several pathways jointly promote PT growth leading to polarized tip growth. Vesicle-mediated transport facilitates the cellular processes required to produce enough cell wall and membrane material for the rapid extension of the pollen tube. Actin dynamics is crucial for maintaining the shape of the PT, and tip growth. Apical secretion controlled by GTPase activity and  $\text{Ca}^{2+}$  gradient is the prime regulators of the secretory pathway, including the apical exo- and endocytosis.

The PT is an outgrowth of the fully mature hydrated pollen grain upon its arrival on the stigmatic surface. It delivers the male gametes to the female gametophyte to ensure double fertilization (Esau, 1977; McCormick, 1993). Once the pollen grain is fully mature, it is released from the anther in a dehydrated state, containing two sperm nuclei and a vegetative cell enclosed within two layers: the inner layer called “intine” and the outer layer is known as “exine” (Figure 1.3 b). The intine is primarily composed of cellulose while the exine is rich in sporopollenin, a highly stable, polymer containing long-chain fatty acids and phenolics (Ariizumi and Toriyama, 2011). The pollen grains of *Arabidopsis* represent individual organisms, containing all the necessary transcripts and proteins required for their further development upon pollination.

During pollination the dehydrated pollen comes into contact with the stigmatic papillar cells (SPC), which upon recognition allow hydration of the pollen grain. The hydrated pollen grain then starts to germinate on the stigmatic surface, which leads to PT emergence that penetrates the stigma and elongates towards the ovule via the transmitting tract to deliver the two sperm cells to the FG. Pollen recognition, germination, and directed growth into the style are controlled by the long chain fatty acids in the pollen outer coat (Hiscock and Allen, 2008; Zinkl *et al.*, 1999; Preuss, 1995). In many species pollen germination is controlled by a sporophytic self-incompatibility (SI) recognition system, which rejects self-pollen and accepts non-self-pollen (Dickinson, 1995; Shiba, 2001; Kachroo *et al.*, 2001). The haploid PT represents the only “migrating” cell in angiosperms (Sanders and Lord, 1989) that elongates by polar tip growth through the transmitting tract (TT) for the delivery of two sperm cells to the FG encased within ovule (Lord, 2003). Tip growth of pollen tube within the TT is regulated by various molecular cues, such as a transmitting tissue-specific (TTS) protein in tobacco (*Nicotiana tabacum*) (Cheung *et al.*, 1995) or gama-aminobutyric acid (GABA) in *Arabidopsis* (Wilhelmi and Preuss, 1996). The later was proposed in helping guiding the PT towards the embryo sac. Additionally, PT growth is characterized by two mechanisms: a steep calcium gradient at the PT tip and reverse fountain streaming due to an atypical organization of the cytoskeleton. The PT exhibits a sharp, tip-focused intracellular calcium gradient that drives and orients its unidirectional apical growth. Plant cells in general maintain a cytosolic calcium concentration of approximately 100nM, which during PT growth increases to several micro molar in the apical-most micrometers of the growing tip (see; Holdaway-Clarke and Hepler, 2003; Malho *et al.*, 1994). In nascent lily PTs the apical calcium concentration was estimated to be 3-5 $\mu$ M, while the rest of the tube maintained a normal calcium level of  $\sim$ 100nM. Cytoskeletal reorganization and dynamics facilitate the vesicular trafficking and targeting of vesicles to the apical plasma membrane, which deliver new membrane and cell wall components to the site of PT growth. For example, inhibition of actin dynamics by applying latrunculin B and cytochalasin B effectively

blocks PT elongation (Miller *et al.*, 1999). On the other hand the role of microtubules in PT growth is puzzling, since in different related studies the pharmacological disruption of microtubules caused different phenotypes (see, Astrom *et al.*, 1995; Laitinen *et al.*, 2002; Anderhag *et al.*, 2000). Furthermore, as in other polar growing cells (e.g. neuronal axons, fungal hyphae, and plant root hair) a conserved self-organizing growth system based on molecular switches called Rho-GTPases of plant (ROPs) functions to establish and maintain polar tip growth by controlling the dynamics of actin filament formation, exocytosis, the production of Reactive Oxygen Species (ROS) and  $\text{Ca}^{2+}$  homeostasis (Lin *et al.*, 1996) (reviewed in Qin and Yang, 2011) (Figure 1.3 c).

#### 1.2.1.2 Mutations that affect male fertility

Disruption of genes responsible for pollen development leads to male sterility or semi-sterility and has been described in more than 100 species (Preuss, 1995). Mutations can have an effect either before or after meiosis. If the mutation has occurred in anther development process or before meiosis then no or very little pollen is produced and all the pollen grains within an anther are affected, plants bearing such anthers are usually male-sterile. However, mutants are termed gametophytic if they disrupt genes that act after meiosis, in the haploid phase of pollen development, and thus only the pollen grains carrying the mutant allele are affected and the plant still carry pollen that are male fertile. In *Arabidopsis*, several sporophytic and gametophytic mutations have been identified, few of which have been extensively characterized, such as a transposon-tagged male sterile mutant in *Arabidopsis thaliana* using the maize *Enhancer-Inhibitor* (Aarts *et al.*, 1993) and a mutant with a defect in the salvage pathway (Regan and Moffatt, 1990). Furthermore, *Delayed dehiscence 2-2 (dde2-2)* a sporophytic mutant, identified in an *En1/Spm1* transposon-induced mutant population screened for plants showing defects in fertility. The *dde2-2* mutant is defective in the anther dehiscence process as well as anther filament elongation and, thus, exhibits a male sterile phenotype. *Dde2-2* thus serves as an excellent tool for manual pollination as shown in (Figure 1.4 a, b) (Von Malek *et al.*, 2002). The mutant phenotype of the plant can be restored by application of methyl jasmonate (JA or MeJA). The *dde2-2* mutant showed complete female fertility when crossed with wild-type pollen as shown in (Figure 1.4 c).

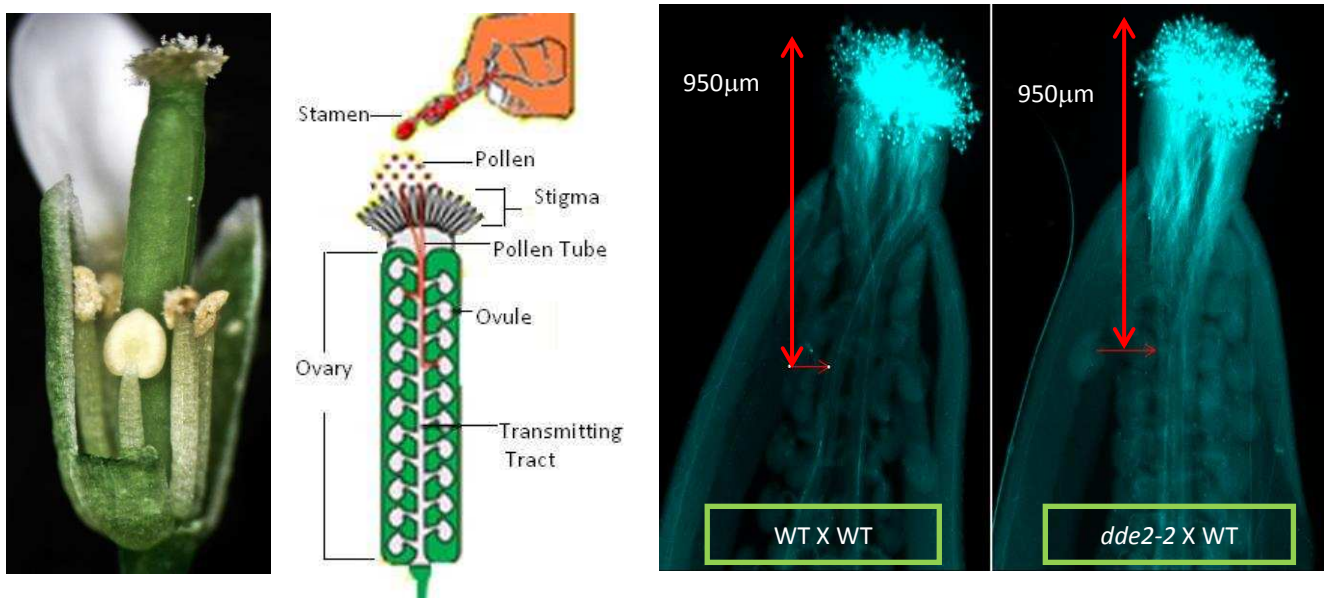
In contrast to the sporophytic *dde2-2* mutant *duo pollen1 (duo1)* is a gametophytic mutant. *DUO1* is male germline-specific transcription factors that coordinates germ cell division, gamete speciation and is essential for sperm cell specification and fertilization. Mutant generative cells in *duo1* pollen fail to enter mitosis at G2-M transition (Durberry *et al.*, 2005).

Recently the role of *ANX1* and *ANX2* (CrRLK1L family members) was explored in PT growth by two separate groups (Boisson-Dernier *et al.*, 2009; Miyazaki *et al.*, 2009). PTs carrying T-DNA insertions in *ANX1* and *ANX2* exhibit normal germination but prematurely rupture failing to fertilize the FGs.



Rare double homozygous mutant display a complete male sterile phenotype (Miyazaki *et al.*, 2009; Boisson-Dernier *et al.*, 2009).

Finally, when the PT grows in the transmitting tract, it responds to various guidance cues from maternal tissue (Johnson and Lord, 2006). The identification of a few sporophytic factors that provide guidance cues at early stages of PT growth have been reported (Dong *et al.*, 2005). However, the signals produced by the female gametophyte itself, which controls the late stages of PT guidance, (Chen *et al.*, 2007; Higashiyama *et al.*, 2001; Márton *et al.*, 2005) were unknown, until the recent identification of the defensin-like LURE polypeptides, which serve as the synergid-secreted PT attractants in *Torenia fournieri* (Okuda *et al.*, 2009). However, how the PT decodes these cues remains unclear as only one male gametophytic mutation called *pollen defective in guidance1* (*POD1*) which disrupts PT guidance, has been identified so far. *POD1* encodes a conserved protein of unknown function and is also essential for positioning and orienting the cell division plane during early embryo development (Li *et al.*, 2011)



**Figure 1.4:** (a) *dde2-2* mutant: *Arabidopsis dde2-2* mutant is a male sterile mutant owing to a frame shift mutation in the AOS gene, which encodes a key enzyme of jasmonic acid (JA) biosynthesis. Therefore, the fertility of the mutant plant can be restored by applying methyl jasmonate. *Dde-2* is defective in anther dehiscence and filament elongation, because of which it is self-infertile and facilitates manual pollination. (b) Picture showing manual pollination (c) Aniline blue staining showing the mutant plant with complete female fertility and normal PT growth when crossed with wild-type pollen.

## 1.2.2 The Arabidopsis Female Gametophyte

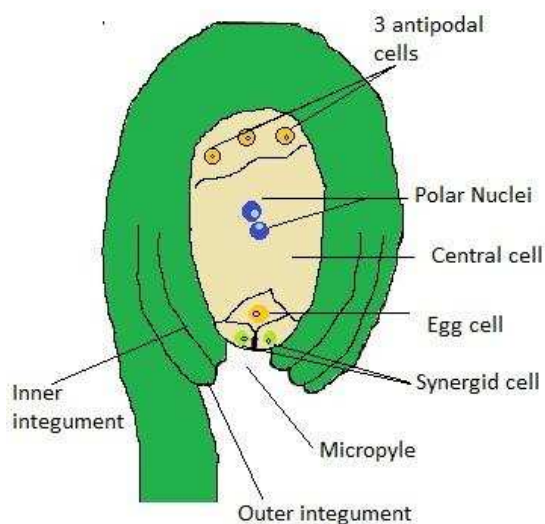
### 1.2.2.1 Development and structure of the female gametophyte

Although both male and female gametophytes of angiosperms are in close association with their respective sporophytic tissue, it is the female gametophyte which is more dependent on the sporophyte because it is encased within the ovule harbored inside the pistil throughout its life cycle. The *Arabidopsis* pistil is composed of three major parts: the stigma, which is the receptive surface and regulates the reception of compatible or incompatible pollen grains; a relatively short



style; and the ovary, which encloses approximately 50 ovules per pistil. The stigmas of angiosperms can be broadly classified into two categories *i.e.* wet or dry, based on their surface secretion. In all wet-stigma species such as *Nicotiana*, *Petunia* (*Solanaceae*) and *Lilium* (*Liliaceae*) the pollen capture and adhesion by stigmatic secretions appears to be non-specific and the hydration of pollen is largely unregulated (Heslop-Harrison & Shivana, 1977; Hiscock and Allen, 2008). In contrast, the studies conducted on dry-stigma species, mostly from *Brassicaceae* (e.g. *Arabidopsis*), suggest high species-specificity for pollen adhesion and capture, with the hydration of pollen on the stigmatic surfaces being tightly regulated (Dickinson, 1995; Hiscock and Allen, 2008).

The sporophytic tissue of reproductive organs produces two types of spores; microspores and megaspores which lead to the development of the male and female gametophytes, respectively. The sporophytic enclosure protects the ovule and the developing seeds from mechanical injury. Within each ovule there is a single embryo sac or megagametophyte that contains six mononucleate cells (three antipodal cells, two synergid cells, and one egg cell) and a binucleate central cell, all encased within a common cell wall, as shown in (Figure 1.6).



**Figure 1.6:** Female gametophyte: The FG consists of seven cells: one large central cell (CC), three antipodal cell (AC), two synergid cells (SC) and one egg cell (EC). The embryo sac is encased within two integuments, an outer and inner. The FG is harbored inside the ovule, which is attached to the placenta by the funiculus.

In angiosperms, reproductive development occurs over two phases referred to as megasporogenesis and megagametogenesis. During megasporogenesis, the diploid megaspore mother cell undergoes meiosis and produces four haploid cells; three of which degenerate and only one survives. During megagametogenesis, the functional megaspore gives rise to the mature female gametophyte. First, the megaspore undergoes several rounds of mitosis without cytokinesis, resulting in a multinucleate coenocyte followed by cellularization around these nuclei, which results in the formation of the female gametophyte. For example in *Arabidopsis*, FG development follows a Polygonium-type pattern, in which the nucleus undergoes two rounds of mitosis, producing a four-nucleate cell having two nuclei at each pole. During the third and final set of mitosis, a cell plate is formed between the sister and non-sister nuclei and, soon thereafter,

the FG cells become completely surrounded by cell membranes. During cellularization, two nuclei, one from each pole (polar nuclei), migrate towards the center of the developing FG and fuse, either before or upon fertilization of the central cell. These series of events results in a seven-celled structure consisting of three antipodal cells, two synergids cells, one central cell and one egg cell (Yadegari and Drews, 2004).

#### 1.2.2.2 1.2.3.2 Mutations that affects female fertility

As angiosperms share an alternation of sporophytic and gametophytic generation, they can be affected by two types of mutations associated with each generation. Sporophytic mutants affect the diploid sporophyte, including megaspore mother cell development, meiosis, and the control of FG development by the surrounding sporophytic tissue. Such mutations follow a Mendelian 1:2:1 segregation pattern. On the contrary, gametophytic mutants do not follow a Mendelian segregation pattern and can be transmitted from generation to generation only as heterozygotes. They affect FG development after meiosis, including megagametogenesis and the functioning of the mature FG (PT guidance, PT reception, fertilization, seed development or maternal control of seed development). Gametophytic mutations can be identified either by segregation distortion and/or reduced seed set. A reduction in seed by 50% set results from a plant heterozygous for a fully penetrant FG mutation, as half of the FGs are mutant and non-functional; thus, they fail to initiate normal seed development. Segregation distortion occurs because, gametophytic mutations are transmitted to subsequent generations at a reduced frequency (Page and Ueli Grossniklaus, 2002; Moore *et al.*, 1997). More than 150 female gametophytic mutants have been reported to date. These include *feronia* (a CrRLK1L member (Escobar-Restrepo *et al.*, 2007)), *scylla* (unknown; Rotman *et al.*, 2003), *lorelai* (*lre*; a putative glycosylphosphatidylinositol (GPI)-anchored protein) (Capron *et al.*, 2008), and *nortia* (*nta*; a mildew resistance locus o (*MLO*) family member (Kessler *et al.*, 2010), which affect PT reception and exhibit a PT over-growth phenotype inside the ovule.

## 1.3 Pollination

### 1.3.1 General observations

Reproduction in plants is initiated immediately after pollen and stigma come into contact during pollination, when pollen is deposited on the stigma by pollinators, wind or through self-pollination. This process is also known as “pollen-stigma interaction” which involves cell-cell recognition processes and is often mediated by a self-incompatibility system, which discriminates self and non-self-pollen and rejects or accepts it, respectively.

In flowering plants, whether self-compatible (SC) or self-incompatible (SI), the pollen grains upon maturation are dispersed from anther and carried to the stigma by pollinators, wind, or direct physical contact, as it is the case in self-fertile plants like *Arabidopsis*. Depending on the species, the pollen may land and germinate on wet or dry stigmatic surface (Heslop-Harrison and Shivanna, 1977). Wet stigmas are, however, at an apparent disadvantage because their wet surface at times permits the capture and development of the pollen of other species and the spores of the pathogens. However, the style of some wet species secretes some enzymes such as glucanases and chitinases, which act as a defence against pathogens. The partial development of the inter-specific pollen on stigmatic surface depletes female resources and physically obstructs the passage of compatible PTs. The situation in wet stigmas is further cofounded by the SI system operating in these plants, which utilizes SI-specific RNase to reject self PTs within the style (Dickinson, 1995), rather than on the stigmatic surface. Dry stigmas thus seem to have an advantage over wet stigmas and can reasonably be regarded as being evolutionarily more advanced. This efficiency is also extended to the SI recognition system for example, if cross-pollen and self-pollen are placed next to each other on a dry stigma surface; the cross-pollen develops a tube, while the self-pollen is immediately arrested. The SI system is thus the best defined example of inter-cellular communications in plants. A wealth of information is available on the development and rejection of non-self and self-pollen on the dry stigmas of Brassicacea (Samuel *et al.*, 2011; Schopfer, 1999; Shiba, 2001; Heizmann *et al.*, 2000).

### 1.3.2 Pollen-stigma interaction in compatible pollinations

In the case of a compatible pollination, the pollen grain lands on the stigmatic surface in a dehydrated state. Immediately after deposition the pollen grain adheres to the stigmatic surface, requiring approximately 10-15 minutes for successful adhesion. This is followed by the recognition event accomplished by the stigmatic papillar cells. As the pollen grain is recognized, the pollen coat forms a junction with the papillar cell, creating a hydraulic contact between the two partners. The pollen coat is relatively impermeable to water, but as a result of the interaction with the stigmatic surface, the pollen coat reorganizes itself both structurally and physiologically (Elleman and Dickinson *et al.*, 1986), and becomes permeable to water. This leads to exchange of water and nutrients required for the hydration and germination of pollen grains. The pollen of “decoated” mutants fail to adhere and hydrate upon contact with the stigma, which shows that the pollen coat plays an integral role in pollen-stigma interactions (Doughty *et al.*, 1993). A long time ago, it was observed that the number of pollen grains on the stigma surface affects the germination rate of pollen. This effect is known as “mentor effect” and only recently, it was found that a secreted peptide called “phytosulphokine” induces germination in pollen population of low numbers (Chen

*et al.*, 2000). Upon hydration, the pollen grain initiates its germination, leading to the emergence of the PT. Once the PT starts to grow, it must penetrate the cuticle and cell wall of the stigmatic papillar cells and enter the style. In dry stigma types, the adhesion of the pollen grain to the stigmatic surface determines the point of entry of the PT into the papillar cells. However, the situation in wet stigmas is different: the extracellular matrices (ECMs) which cover the stigma, govern PT growth till PT has a certain length before the PTs finally enter the stylar tissue. Upon emergence at the base of the stigmatic papillar cells, the PT grows intercellular down the style and then through the transmitting tract of the septum (a central tissue that runs to the base of the ovary). The time and distance travelled in the style is relatively short in comparison to that spent in the transmitting tissue. Soon after the PT leaves the transmitting tract, it starts to grow towards the funiculus. During its growth, the heterotropic PT continues to absorb the required nutrients from the female tissues. Soon, the PT reaches the ovule and then starts to grow along the funicular surface to enter micropyle opening. Soon upon PT arrival at the micropyle, the PT enters one synergid cell where it bursts and releases the two sperm cells. These fuse with the egg and central cell to form the diploid zygote and the triploid endosperm, respectively.

The growth and guidance of the pollen tube in the transmitting tract is facilitated by a proposed chemo-attractive gradient present in the pistil, which plays an important role in guiding the pollen tube towards the ovule (Lord, 2003). In *Arabidopsis* one of the first molecules proposed to guide PTs was  $\gamma$ -amino butyric acid (GABA), which was identified through the analysis of the *Arabidopsis pollen-pistil 2 (pop2)* mutant (Wilhelmi and Preuss, 1996). Further advances in the understanding of chemo-attractants were made by the discovery of chemocyanin in lily (Kim *et al.*, 2003) which is structurally similar to the plantacyanin in *Arabidopsis* (Dong *et al.*, 2005). Both these peptides are present in an increasing gradient along the style and transmission tract leading to the ovary.

### 1.3.3 Pollen-stigma interactions in self-incompatible species

A lot of work has been done on pollen-stigma interactions of SI species of angiosperms (Heslop-Harrison, 1978). No review on pollen-stigma interactions would conclude without at least a brief discussion of SI and its relationship with the events that define compatible pollen-stigma interactions.

SI is a genetically determined pre-zygotic barrier to self-fertilization and fertilization between individuals with identical haplotypes of the polymorphic Sterility (*S*) locus. Gametophytic self-incompatibility (GSI) and sporophytic self-incompatibility (SSI) are two genetically distinct forms of SI. In GSI, the *S* phenotype of pollen is determined by its own haploid genome, whereas in SSI, the *S* phenotype of the pollen is determined by the diploid phenotype of the parent plant. This suggests that a dominant interaction between *S* haplotypes is possible in SSI but not in GSI

(reviewed in, Hiscock and Allen, 2008). GSI and SSI differ from each other in the timing of their incompatibility reaction. GSI is mainly associated with the inhibition of incompatible PTs within the style; however, in SSI, incompatible pollen is usually inhibited during the pollen–stigma interaction, long before PT emergence. There are exceptions to this rule as seen in poppy (*Papaveraceae*) and grasses (*Poaceae*), both of which have GSI and yet inhibit the incompatible pollen at the stigmatic surfaces. In the Brassica SSI system incompatible pollen does usually not hydrate but can be arrested at any stage during the pollen–stigma interaction, from pre-hydration to post-penetration of the papillar cell. However, the PT rarely penetrates the stigmatic wall and the majority of incompatible pollen grains cease development either before or just after germination (Dickinson *et al.*, 1986). Incompatible pollen is recognized by a S-haplotype-specific protein-protein interaction between a stigma-specific S-receptor serine/threonine kinase, SRK (female determinant of SI) and its cognate ligand, the S cysteine-rich peptide, SCR (male determinant) (Kachroo *et al.*, 2001). Expression of *SRK* occurs in the papillar cells of the stigma and is localized in the plasma membrane. The SCR interaction with SRK triggers a kinase-mediated signalling cascade that arrests the pollen and blocks pollination. In *Arabidopsis* both SRK and SCR are pseudo genes allowing self-pollination (Kusaba *et al.*, 2001).

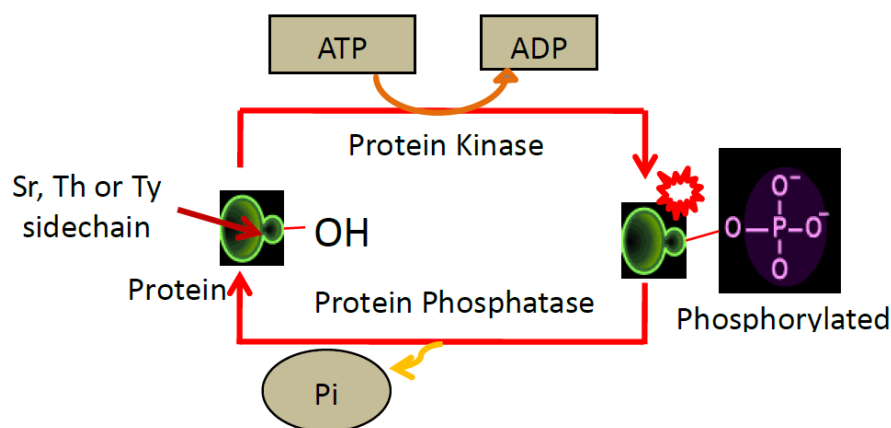
## 1.4 Post Translational Modifications: Phosphorylation

### 1.4.1 History, discovery and importance

Within the last few decades, it has been noticed that a proteome of an organism is vastly more complex than its respective genome. For example, the human genome contains 20,000 – 33,000 genes that encode for more than a million proteins. These correlations demonstrate that a single gene encodes multiple proteins; this increase in complexity from the level of the genome to the proteome is further enhanced by post-translational modifications of protein. Post-translational modifications (PTMs) are chemical modifications of proteins after their synthesis, regulating activity, localization and interactions of proteins with other proteins, nucleic acids, lipids, and cofactors. PTMs can modify the proteins at any stage of their life cycle depending on their nature and functional attributes. For example, many proteins are modified shortly after translation is completed to mediate proper protein folding and stability, or to direct the nascent protein to distinct cellular compartments (e.g., nucleus or membrane). Other modifications occur after folding and localization are completed to activate or inactivate the catalytic activity of the protein. Proteins are also covalently marked by tags that target a protein for degradation. Besides single modifications, proteins are often modified through a combination of post-translational cleavage and the addition of functional groups through a step-wise mechanism of protein maturation or

activation. Post-translational modifications can modify the proteins via phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, and lipidation, and can influence almost all aspects of normal cell biology and pathogenesis. Therefore, identifying and understanding PTMs is critical for understanding cell biology and the treatment and prevention of disease.

Phosphorylation was first identified in the protein Vitellin (protein found in egg yolk) in the year 1906, by Phoebus Levene at the Rockefeller Institute for Medical Research. Later in 1932, Phoebus Levene along with Fritz A. Lipmann detected phosphoserine in Casein. Thereafter, phosphorylation turned out to be one of the most biologically relevant and ubiquitous post-translational protein modifications controlling cell division, signal transduction, enzymatic activity, and so forth. Phosphorylation is a reversible process mediated by protein kinases and phosphatases as shown in (Figure 1.7) (reviewed in Paradela and Albar, 2008). Kinases phosphorylate a protein in the presence of ATP and phosphatases dephosphorylate a protein by releasing a phosphate group from the protein.



**Figure 1.7:** Schematic illustrating the reversible phosphorylation reaction. The protein kinase phosphorylates protein in the presence of ATP and the protein phosphatase dephosphorylates it by removing a phosphate group.

#### 1.4.2 Functions, type of phosphorylation and identification

In both prokaryotes and eukaryotes, phosphorylation acts as an important regulatory mechanism. Reversible phosphorylation causes many conformational changes in the protein structure of many enzymes and receptors, thereby regulating their activity by either switching it on or off via phosphorylation or dephosphorylation. In eukaryotes, protein phosphorylation usually occurs on serine, threonine, tyrosine and histidine residues of amino acid sequences. Interestingly, histidine phosphorylation of eukaryotic proteins appears to be more frequent than tyrosine phosphorylation (Cieřla *et al.*, 2011). On the other hand, in prokaryotic proteins, the phosphorylation of serine, threonine or tyrosine are more common, but histidine, arginine or lysine can also be targets of phosphorylation (Cozzone, 1988; Cieřla *et al.*, 2011).

Phosphorylation can control protein–protein interactions, protein–membrane interactions, protein subcellular localization, and protein stability in response to external or internal stimuli (Mithoe and Menke, 2011). As individual phosphoresidues can have distinct functions, it is not sufficient to know whether a protein is phosphorylated or not: rather, it is important to know which particular residue, *i.e.* serine (Ser), threonine (Thr), histidine (His) or tyrosine (Tyr), is phosphorylated in a given phosphopeptide. *In vivo* phosphorylation sites are the direct result of the activity of a specific kinase, and in the case of kinases themselves, the presence of a phosphorylated site (phosphosite) may reflect activation or inactivation of the kinase. Moreover, identifying phosphosites can allow site-directed mutagenesis studies to explore the function of particular proteins. Various methods have been proposed to extract phosphorylation motifs from large sets of phosphorylated peptides. Motif-X (Schwartz and Gygi, 2005) extracts significantly over-represented motifs from large-scale phosphoproteomics data, and thus provides valuable information on the kinase–substrate specificity. In eukaryotes, Ser, Thr and Tyr are the most common targets of phosphorylation. Despite the fact that Tyr phosphorylation is less abundant in eukaryotes as compared to Ser/Thr phosphorylation, it is crucial for protein–protein interactions (Manning *et al.*, 2002; Dissmeyer and Schnittger, 2011). The human genome encodes for approximately 500 kinases and 100 phosphatases, accounting for almost 2.0% of the genome (Manning *et al.*, 2002). The *Arabidopsis* genome encodes approximately 1100 kinases and 100–200 phosphatases, accounting for 5% of its genome, and has a similar number of genes as the human genome (*Arabidopsis* Genome Initiative, 2000). In general, more kinases than phosphatases may be required as protein kinases play a crucial and specific role in the response to a variety of external and internal stimuli, while phosphatases merely reverse the effect of kinases (Smith and Walker, 1996). The vast majority of plant kinases phosphorylate Ser and Thr; only recently Tyr kinases and their role in signalling events during plant development have been discovered (Luan, 2002; Sugiyama *et al.*, 2008; Jaillais *et al.*, 2011).

### 1.4.3 Role of Phosphorylation in pollination

As discussed above, phosphorylation mediates various cellular processes and controls protein–protein interactions, protein–membrane interactions, protein subcellular localization, and protein stability in response to external or internal stimuli. Therefore, in successful pollination, which depends on cell-cell communication between the two reproductive partners, rapid cellular response should occur. In *Arabidopsis*, the mature dehydrated pollen grain upon landing on a dry stigma, hydrates, germinates and grows a PT that delivers the sperm cells to the FG to affect double fertilization. Various studies have emphasized that a mature, dehydrated pollen grain contains all the transcripts and proteins required for germination and initial PT growth (Honys and

Twell, 2004; Borges *et al.*, 2008; Holmes-Davis *et al.*, 2005) and it is assumed that these initial events of pollination are most likely dependent on PTMs, specifically phosphorylation (Holmes-Davis *et al.*, 2005; Grobei *et al.*, 2009). Indeed, it was found that the Gene Ontology (GO) function terms “protein kinase” and “protein phosphatase” are over-represented in the pollen grain transcriptome as compared to transcriptome of leaves, seedlings, and siliques (Pina *et al.*, 2005). Therefore, it is important to explore the role of phosphorylation through which many processes induced by pollination might be controlled.

#### 1.4.4 Phosphoproteomics: Detection and Characterization

Identification of the phosphorylated proteins and understanding the dynamics of protein phosphorylation, including which residues are phosphorylated and dephosphorylated, in response to physiological, cellular and environmental factors, can help in dissecting the regulatory biological networks at a global level. This area of systems biology is termed “phosphoproteomics”. Phosphoproteomics deals with the detection of phosphoproteins from within a complex protein sample but it is often not easy to capture phosphoproteins owing to their low abundance in any given tissue. Approximately 30% of the whole proteome of an organism is phosphorylated, but it is hard to detect them without a proper enrichment technique. Today, advanced technologies which can enrich, detect and quantify the phosphopeptides/phosphoproteins are available and are relatively fast and have a high accuracy (Kersten *et al.*, 2006). This advancement has made the large-scale study of phosphoproteins feasible. Most of these technologies are applicable to plants (Reiland *et al.*, 2009; Mayank *et al.*, 2012; Nühse *et al.*, 2004; Peck, 2006). The establishment of economical, stable and reproducible tools for studying phosphorylation in plants will further boost research in the field of phosphoproteomics.

In today’s world we have a handful of strategies for the enrichment and detection of phosphopeptides/ phosphoproteins in plants which have proven to be beneficial for the understanding of the biology of plants. In the following section I focus on optimal enrichment technique for phosphoproteins/ phosphopeptides.

##### 1.4.4.1 Pre-treatment of the sample

Protein phosphorylation is a highly dynamic process controlled by the reversible activity of protein kinases and protein phosphatases. To capture a phosphorylated protein in a given tissue or cell type it is important to inhibit the activity of the enzymes that dephosphorylate proteins in the sample preparation step. Some of the known compounds that can efficiently inhibit the activity of phosphatases are vanadate (a potent inhibitor of tyrosine phosphatases), ocadaic acid (an inhibitor of serine and threonine phosphatases), sodium fluoride and sodium pyrophosphate (broad-acting or non-specific reagents). Along with the phosphatase inhibitor, it is also advised to



include inhibitors of proteases that conduct proteolysis. There are six different types of proteases: serine proteases, threonine proteases, cysteine proteases, aspartate proteases and metalloproteases (Schmidt *et al.*, 2007), which can be inhibited by variety of compounds

#### 1.4.4.2 Pre-fractionation of the sample based on the charge state: Ion exchange

Differences in the net charge of molecules can be exploited to pre-fractionate the protein of interest from a crude and complex sample. Ion exchange chromatography is used to achieve separation of molecules based on their charge state and the method works efficiently for large protein molecules, nucleotides, peptides and even amino acids. In a chromatography column the injected solution represents the “sample” and individually separated components are termed “analytes”. This type of chromatography is further sub-divided into two types: (1) Strong Cation Exchange; SCX: retains positively charged cations because the stationary phase displays a negatively charged functional group, (2) Strong Anion Exchange; SAX: retains negatively charged anions because the stationary phase displays a positively charged functional group.

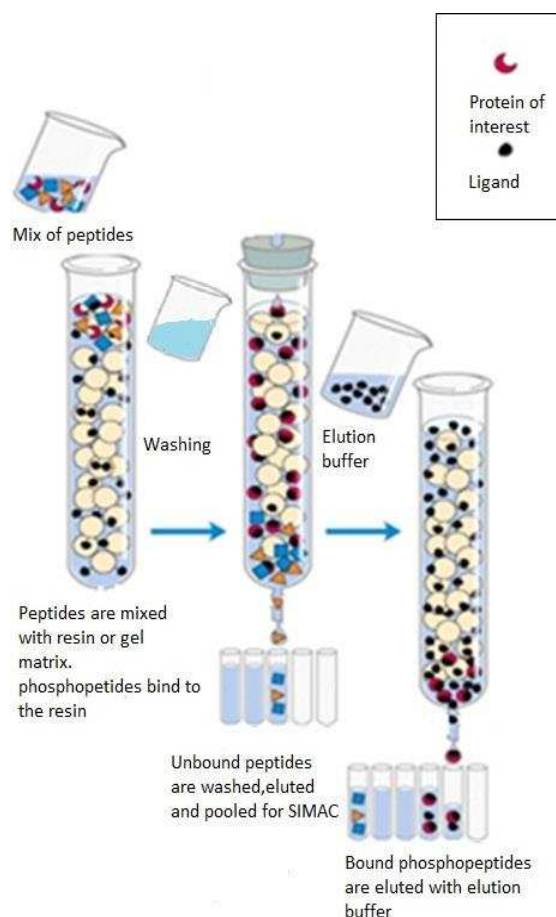
For ion exchange chromatography, a sample is introduced, either using an auto sampler or manually, into a sample injection needle. A buffered aqueous solution, known as the mobile phase, carries the sample from the injection needle to the column that contains the stationary phase, typically a resin or gel matrix consisting of agarose or cellulose beads with covalently bonded charged functional groups. The target analytes either bind or elute from the stationary phase, based on the charge state of the analyte and the column matrix. For example, in SCX when we inject a sample containing phosphopeptides, which are highly negatively charged, they do not bind to the column and rather elute in the initial unbound fraction. However, if the same sample is loaded onto a SAX column, then it will bind to the matrix and the elution of the sample from the column will be performed using similarly charged species in the elution buffer, which will displace the analyte ion from the stationary phase (Leitner *et al.*, 2011).

#### 1.4.4.3 Phosphoprotein and phosphopeptide enrichment techniques

Enrichment of phosphoproteins or phosphopeptides is a crucial step as only 30% or less of any organism’s proteome contains phosphoproteins at a given time. Moreover, these phosphoproteins are often expressed at very low abundance in the cell, especially the ones with regulatory functions such as kinases and phosphatases involved in signal transduction. Nowadays, commercial kits for phosphoprotein and phosphopeptide enrichment are available, and can significantly reduce the sample enrichment time and labor, while increasing the quality. Amongst the various enrichment techniques discussed, I focus specifically on phosphopeptide/phosphoprotein enrichment chromatographies:

**IMAC:** Immobilized Metal Affinity Chromatography, was originally developed for the purification of His-tagged proteins, but later it became the most widely used technique for phosphopeptide enrichment (Kersten *et al.*, 2006). Phosphorylated proteins or peptides are bound to the IMAC stationary phase by electrostatic interactions, with negatively charged phosphate groups binding to positively charged metal ions, which in turn are bound to the column material itself via Tris (carboxymethyl) ethylenediamine (TED) iminodiacetic acid (IDA), and nitriloacetic acid (NTA) linkers. Initially, metal ions such as  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Mn}^{2+}$  were shown to have a higher affinity for proteins having a high density of histidine; however, immobilized metal ions such as  $\text{Fe}^{3+}$ ,  $\text{Ga}^{3+}$ , and  $\text{Al}^{3+}$  have been shown to have a greater affinity for phosphopeptides. Recently,  $\text{Zr}^{4+}$  has been shown to bind phosphopeptides with a high affinity and specificity (Kweon and Håkansson, 2006). One of the major drawbacks of IMAC chromatography is the non-specific binding of highly acidic amino acids such as Glu and Asp, and the strong binding of multiply phosphorylated peptides, which hinders the binding of mono-phosphorylated peptides. Despite IMAC's easy protocol (binding-washing-eluting) (Figure 1.8); its experimental conditions are variable, with even a small variation in pH or ionic strength of the solvent possibly leading to non-reproducible results. IMAC protocols have become very popular, r due to their good compatibility with subsequent detection techniques such as LC-ESI-MS/MS or MALDI MS (Garcia *et al.*, 2004).

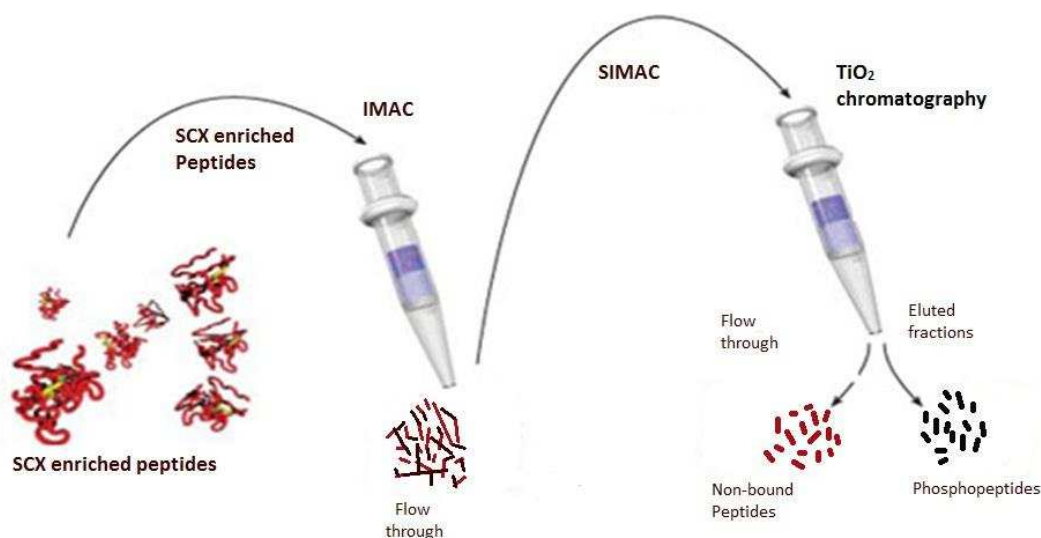
**$\text{TiO}_2$ :** titanium dioxide chromatography is a strong and promising alternative to IMAC for the enrichment of phosphorylated peptides. This chromatography was first described by Pinkse and colleagues (Pinkse *et al.*, 2004). Their protocol was optimized to enrich phosphopeptides, but could not prevent the non-specific binding of acidic non-phosphorylated peptides. Larsen and colleagues later improved the selectivity of phosphopeptides using  $\text{TiO}_2$  (Larsen *et al.*, 2005). They loaded the peptides onto the column using dihydroxy-benzoic acid (DHB). DHB proved to be very efficient in reducing the binding of non-phosphorylated peptides to  $\text{TiO}_2$ , while retaining high binding affinity for phosphorylated peptides. In a recent study Bodenmiller and coworkers described provided optimal enrichment of phosphorylated peptides loaded with phthalic acid onto the  $\text{TiO}_2$  column (Bodenmiller *et al.*, 2007). Furthermore, non- specific binding can also be greatly reduced by adding a high concentration of trifluoroacetic acid (TFA) in the sample buffer.  $\text{TiO}_2$  has an advantage over IMAC because it is fast (typically 5 mins per sample) and can also be coupled with detection techniques such as LC-ESI-MS/MS.



**Figure 1.8:** Demonstrating the phosphopeptides enrichment by IMAC or  $\text{TiO}_2$  chromatography.

**SIMAC:** Sequential Elution from IMAC, provides a sequential separation of mono-phosphorylated peptides from a complex protein mixture. Mono-phosphorylated peptides are not retained in IMAC, because of its stronger affinity for multiply phosphorylated peptides. There are mainly two ways of conducting SIMAC. Firstly, using an acidic elution in IMAC chromatography, followed by  $\text{TiO}_2$  chromatography, this holds the mono-phosphorylated peptides, whereas a subsequent basic elution recovers the multiply phosphorylated peptides. The second way is by conducting  $\text{TiO}_2$  chromatography on the unbound fraction of IMAC, which contains all the unbound mono-phosphorylated peptides as shown in (Figure 1.9). Overall, SIMAC provides an additional enrichment step which helps to retain the mono-phosphorylated peptides that are not retained in either IMAC or  $\text{TiO}_2$  chromatography (Thingholm *et al.*, 2008).

The enrichment efficiency of all three chromatographies is discussed in more detail in Chapter 3.



**Figure 1.9:** SIMAC: The mono-phosphorylated peptides, which are not retained in IMAC chromatography, are present in the flow through. This flow through is further enriched with  $\text{TiO}_2$  chromatography to retain all the mono-phosphorylated peptides. Most of these peptides are unique and the overlap with IMAC and  $\text{TiO}_2$  is low (discussed in more detail in chapter 3)

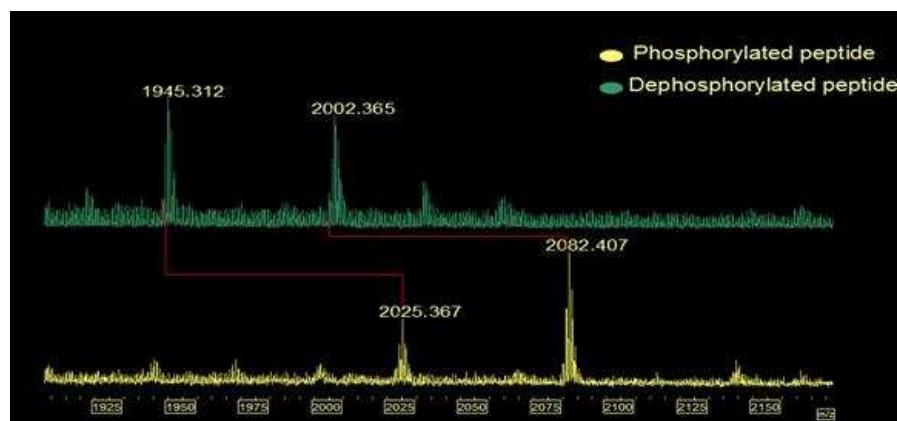
#### 1.4.5 Detection of enriched phosphopeptides: Mass spectrometry

The phosphopeptides are negatively charged and, therefore, can be measured by a mass-spectrometer (MS) that can determine the mass of the molecule based on its charge state. MS is an analytical technique that can also determine the element composition of a sample in the molecule, as well as elucidate the chemical structures of molecules, such as peptides and other molecules. MS ionizes the chemical compounds to generate charged molecules or fragments, and measures their mass-to-charge ratio. In the late 1980s, technical advancements in MS revolutionized protein chemistry and the analysis of proteins with the launch of two ionization techniques: Electro Spray Ionization (ESI) and Matrix-Assisted Laser Desorption/Ionization (MALDI) (Aebersold and Goodlett, 2001). These ionization techniques reduced the problem of generating ions from large non-volatile analytes, such as proteins and peptides, without causing excessive fragmentation of these molecules. Therefore, they are also known as “soft” ionization methods. ESI gained more popularity because of the ease with which it can interface with popular chromatographic and electrophoretic liquid-phase separation techniques and it quickly supplanted fast atom bombardment as the ionization method of choice for protein and peptide samples in liquid phase. Moreover, MALDI was later recognized for its time-of-flight (TOF) analyser, which is more robust, simple and sensitive, and has a larger mass range. MALDI mass spectra are also relatively easy to interpret because they predominantly generate singly charged ions. MALDI makes use of the impact of high energy photons to vaporise and ionise a sample embedded in a solid organic matrix and spotted on a metallic plate, whereas ESI is used to produce ionized

molecules in the gas phase from a liquid injection by creating a spray of droplets in the presence of a strong electric field.

A mass spectrometer has three major components: the ionization source (ESI or MALDI), the mass analyser (where the molecule is separated on the basis of mass-to-charge ratio in the presence of an electromagnetic field; such as TOF, quadrupole, ion trap, FTICR), and detector (provides data for calculating the abundance of each ion present in the injected sample). Many mass spectrometers use two or more mass analyzers for tandem mass spectrometry (MS/MS). A tandem mass spectrometer is capable of multiple rounds of MS, usually separated by some form of molecule fragmentation. For example, one mass analyzer can isolate one peptide out of many entering the mass spectrometer. A second mass analyzer then stabilizes the peptide ions while they collide with a gas, causing them to fragment by either collision-induced dissociation (CID) or by electron transfer dissociation (ETD). Sometimes ETD is combined with CID, to form a third mass analyzer, which further sorts the fragments produced from the peptides.

The output file from a mass spectrometer is called “mass spectrum” and it plots the ion intensity vs  $M/z$  (mass-to-charge ratio) (Gross, 2004) (Figure 1.10). The mass spectrum of a phosphorylated peptide differs from a non-phosphorylated peptide by a mass shift of 80 Dalton as shown in the figure below.



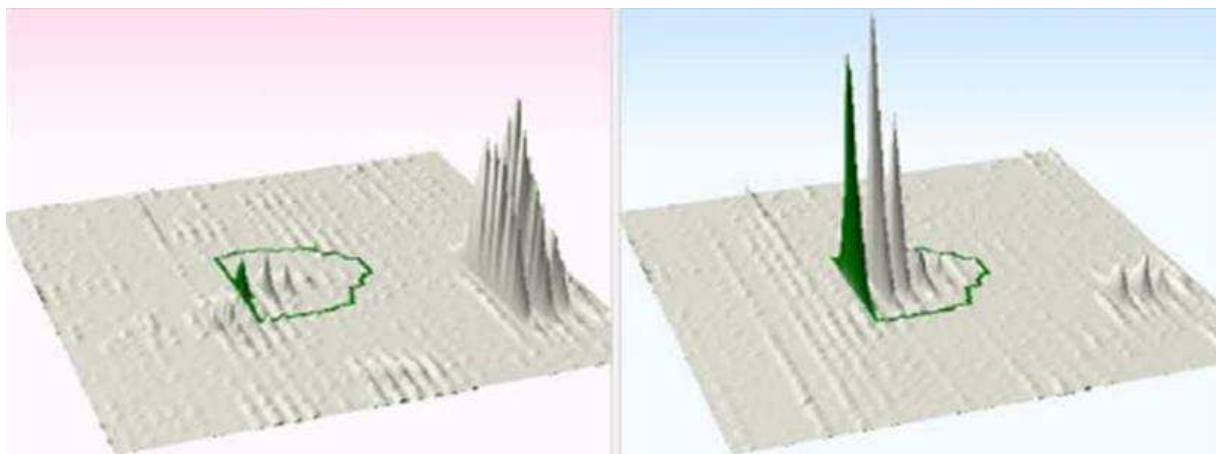
**Figure 1.10:** Mass spectrum: Showing two intense phosphorylated peaks with masses 2082.407 and 2025.387 Dalton, with their dephosphorylated masses of 2002.365 and 1945.312 respectively. The MS technique can be applied for both qualitative and quantitative analyses and is broadly used to identify unknown compounds, to determine isotopic composition of an element in a molecule, and to determine the sequence of a molecule (protein/peptide) by observing the fragmentation pattern.

#### 1.4.6 Quantification of Identified Phosphopeptides:

MS plays a vital role for the identification and characterization of molecules (proteins, peptides, and metabolites) in a given tissue. In addition to the global profiling of molecules at any time, information about the expression level is increasingly required in biology. This is where quantitative proteomics comes to play and provides either an absolute or relative quantification of proteins/peptides in a given biological sample. It is often not easy to quantify a molecule in a biological sample because of its own complexity, the sensitivity of the instrument, and the low-to-

moderate throughput of the system from sample preparation to data analysis. The most commonly used proteomic analysis approach is achieved by a combination of either two-dimensional gel electrophoresis (2DE) or liquid chromatography (LC), which helps to separate the proteins/peptides with MS in order to visualize, identify, characterize, and quantify them.

Protein/Peptides quantification by LC-MS can be achieved either by a chemical labeling approach or by a label-free approach. The first approach involves incorporation of stable isotopes into one or more samples, which is usually done *in vivo* by stable isotope labeling by amino acids in cell culture (SILAC) method and is introduced in the cell culture (Ong *et al.*, 2002; Ong *et al.*, 2004), or occasionally *in vitro* by chemical or enzymatic means (Cagney and Emili, 2002; Ross *et al.*, 2004). The labeled peptides will then be heavier and are simultaneously analyzed along with the control peptides, thereby allowing them to be distinguished in the mass spectrometer. The labeled approach has a limitation with samples sets (limited amount of samples can be analyzed in parallel) and is very expensive for routine usage. However, the second approach, which is label-free, is more feasible for comparative quantification for large sample sets (no limit to samples). This type of approach is simple and cost-effective for routine usage on any sample type. The label-free approach was shown to have a high reproducibility and linearity at both the peptide and protein levels. The limiting factor for this strategy is the sample preparation step, as it is advised to prepare all the samples meant to be quantified in parallel using identical reagents, preferably at similar environmental conditions to avoid technical variations. PROGENESIS-LC MS/MS is a web based tool which can be locally downloaded on a private computer that identifies and quantifies the proteins/peptides without chemical labeling of the sample. Therefore, it provides a label-free relative quantification of proteins in a given sample. The principle of label-free quantitative proteomics is based on the comparison of precursor ion intensities across all experiments once all features (defined as isotopic clusters) are aligned according to their LC retention time,  $m/z$  and charge states. There are two crucial requirements in order to make this approach successful. First, the retention time, precursor mass/charge ratio, and charge state are usually the parameters by which a “feature” is defined, and the alignment between different sample sets is achieved, and secondly the requirements for reproducible LC retention time, high resolution and mass accuracy for mass spectrometers are critical to ensure that the same peptides are compared across all experiments. (Figure 1.10) shows the quantification achieved by Progeneis LC-MS, further details are discussed in Chapter 4.



**Figure 1.10:** Progenesis LC-MS 3D view: Showing a 3D view of peak intensity of one representative peptide obtained from Progenesis LC-MS. Adapted from: Functional Genomics Centre Zurich (FGCZ) Website.

## 1.5 References

- Aarts, M., Dirkse, W., Stiekema, W. and Pereira A.,** (1993) Transposon tagging of a male sterility gene in *Arabidopsis*. *Nature.*, 363(6431):715-7.
- Aebersold, R and Goodlett, D.R.,** (2001) Mass spectrometry in proteomics. *Chemical Reviews*, 101(2), 269–95.
- Anderhag, P., Hepler, P. K. and Lazzaro, M.D.,** (2000) Microtubules and microfilaments are both responsible for pollen tube elongation in the conifer *Picea abies* (Norway spruce). *Protoplasma*, 214(3-4), 141–157.
- Ariizumi, T. and Toriyama, K.,** (2011) Genetic regulation of sporopollenin synthesis and pollen exine development. *Annual Review of Plant Biology*, 62, 437–60.
- Astrom, Sorri, O. and Raudaskoski, M.,** (1995) Role of microtubules in the movement of the vegetative nucleus and generative cell in tobacco pollen tubes. *Sexual Plant Reproduction*, 8(2), 61 – 69.
- Bell, G.,** (1994) *The evolution of haploid-diploid life cycles: 1993 Symposium on Some Mathematical Questions in Biology, June 19-23, 1993, Snowbird, Utah*, American Mathematical Soc.
- Bennett, M.D. and Smith, J.B.,** (1976) Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 274(933), 227–74.
- Bodenmiller, B., Mueller, L.N., Mueller, M., Domon, B. and Aebersold, Ruedi,** (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods*, 4(3), 231–7.
- Boisson-Dernier, A., Roy, S., Kritsas, K., Grobei, M. a, Jaciubek, M., Schroeder, J.I. and Grossniklaus, U,** (2009) Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge. *Development*, 136(19), 3279–88.
- Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, Sheila, Feijó, J. a and Becker, J.D.,** (2008) Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiology*, 148(2), 1168–81.



- Cagney, G. and Emili, A.,** (2002) *De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nature Biotechnology*, 20(2), 163–70.
- Capron, A., Gourgues, M., Neiva, L.S., et al.,** (2008) Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene. *The Plant Cell*, 20(11), 3038–49.
- Chen, Y.-H., Li, H.-J., Shi, D.-Q., Yuan, L., Liu, J., Sreenivasan, R., Baskar, R., Grossniklaus, U and Yang, W.C.,** (2007) The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. *The Plant Cell*, 19(11), 3563–77.
- Chen, Y.F., Matsubayashi, Y. and Sakagami, Y.,** (2000) Peptide growth factor phytosulfokine- $\alpha$  contributes to the pollen population effect. *Planta*, 211(5), 752–5.
- Cheung, A.Y., Wang, H. and Wu, H.M.,** (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell*, 82(3), 383–93.
- Cieřła, J., Frączyk, T. and Rode, W.,** (2011) Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochimica Polonica*, 58(2), 137–48.
- Coelho, S.M., Peters, A.F., Charrier, B., Roze, D., Destombe, C., Valero, M. and Cock, J.M.,** (2007) Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene*, 406(1-2), 152–70.
- Cozzone, A.J.,** (1988) Protein phosphorylation in prokaryotes. *Annual Review of Microbiology*, 42, 97–125.
- Dean, C. and Schmidt, R.,** (1995) Plant genomes: A current molecular description. *Annual Review of Plant Physiology and Plant Molecular Biology*, 46(1), 395–418.
- Dickinson, H.,** (1995) Dry stigma, water and self-incompatibility in Brassica. *Sexual Plant Reproduction*, 8(1), 1–10.
- Dickinson, H., Doughty, J., Hiscock, S.J., Elleman, C.J. and Stephenson, a G.,** (1986) Pollen-stigma interactions in Brassica. *Symposia of the Society for Experimental Biology*, 80, 141–157.
- Dissmeyer, N. and Schnittger, A.,** (2011) The age of protein kinases. *Methods in Molecular Biology*, 779, 7–52.

- Dong, J., Kim, S.T. and Lord, Elizabeth M.,** (2005) Plantacyanin plays a role in reproduction in *Arabidopsis*. *Plant Physiology*, 138(2), 778–89.
- Doughty, J., Hedderson, F., McCubbin, A. and Dickinson, H.,** (1993) Interaction between a coating-borne peptide of the Brassica pollen grain and stigmatic S (self-incompatibility)-locus-specific glycoproteins. *Proceeding of the National Academy of Sciences* , 90(2), 467–71.
- Durbarry, A., Vizir, I. and Twell, D.,** (2005) Male germ line development in *Arabidopsis*. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. *Plant Physiology*, 137(1), 297–307.
- Esau, K.,** (1977) Anatomy of seed plants. *New York: John Wiley & Sons*.
- Escobar-Restrepo, J.M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W.C. and Grossniklaus, U.,** (2007) The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science*, 317(5838), 656–60.
- Garcia, B.A., Busby, S.A., Barber, C.M., Shabanowitz, J., Allis, C.D. and Hunt, D.F.,** (2004) Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *Journal of Proteome Research*, 3(6), 1219–27.
- Graham, L.K. and Wilcox, L.W.,** (2000) The origin of alternation of generations in land plants: a focus on matrotrophy and hexose transport. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1398), 757–66; discussion 766–7.
- Grobei, M.A., Qeli, E., Brunner, E., Rehrauer, H., Zhang, R., Roschitzki, B., Basler, K., Ahrens, C.H. and Grossniklaus, Ueli,** (2009) Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Res*, 1786–1800.
- Gross, J.H.,** (2004) Mass Spectrometry - A textbook. *Springer - Verlag*.
- Grossniklaus, U and Schneitz, K,** (1998) The molecular and genetic basis of ovule and megagametophyte development. *Seminars in Cell & Developmental Biology*, 9(2), 227–38.
- Heizmann, P., Luu, D.-T. and Dumas, C.,** (2000) The clues to species specificity of pollination among Brassicaceae. *Sexual Plant Reproduction*, 13(3), 157–161.

- Heslop-Harrison, Y. and Shivanna, K.R.,** (1977) The receptive surface of the angiosperm stigma. *Ann. Bot.*, 41(6), 1233–1258.
- Heslop-Harrison.J.,** (1978) Recognition and response in pollen-stigma interaction. *Symp Soc Exp Biol*, 32, 121–38.
- Higashiyama, T, Yabe, S., Sasaki, N, Nishimura, Y., Miyagishima S, Kuroiwa, H. and Kuroiwa, T,** (2001) Pollen tube attraction by the synergid cell. *Science*, 293 (5534), 1480–3.
- Hiscock S.J and Allen, A.M.,** (2008) Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytol*, 179 (2), 286–317.
- Holdaway-Clarke, T.L. and Hepler, Peter K.,** (2003) Control of pollen tube growth: role of ion gradients and fluxes. *New Phytologist*, 159(3), 539–563.
- Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J. and McCormick, Sheila,** (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics*, 5 (18), 4864–84.
- Honys, D. and Twell, D.,** (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology*, 5(11), R85.
- Jaillais, Y., Hothorn, M., Belkhadir, Y., Dabi, T., Nimchuk, Z.L., Meyerowitz, E.M. and Chory, J.,** (2011) Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes & Development*, 25(3), 232–7.
- Johnson, M.A. and Lord, E.,** (2006) Extracellular guidance cues and intracellular signaling pathways that direct pollen tube growth. *Plant Cell Monographs*, 3, 223–242.
- Kachroo, A., Schopfer, C R, Nasrallah, M.E. and Nasrallah, J.B.,** (2001) Allele-specific receptor-ligand interactions in Brassica self-incompatibility. *Science*, 293(5536), 1824–6.
- Kenrick, P.,** (2000) The relationships of vascular plants. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1398), 847–55.
- Kersten, B., Agrawal, G.K., Iwahashi, H. and Rakwal, R.,** (2006) Plant phosphoproteomics: a long road ahead. *Proteomics*, 6(20), 5517–28.

- Kessler, S.A., Shimosato-Asano, H., Keinath, N.F., Wuest, S.E., Ingram, G., Panstruga, R. and Grossniklaus, U.,** (2010) Conserved molecular components for pollen tube reception and fungal invasion. *Science*, 330(6006), 968–71.
- Kim, S., Mollet, J.-C., Dong, J., Zhang, K., Park, S.-Y. and Lord, Elizabeth M,** (2003) Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. *Proceedings of the National Academy of Sciences*, 100(26), 16125–30.
- Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B. and Nasrallah, M.E.,** (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *The Plant Cell*, 13(3), 627–43.
- Kweon, H.K. and Håkansson, K.,** (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Analytical Chemistry*, 78(6), 1743–9.
- Laitinen, E., Nieminen, K., Vihinen, H. and Raudaskoski, M.,** (2002) Movement of generative cell and vegetative nucleus in tobacco pollen tubes is dependent on microtubule cytoskeleton but independent of the synthesis of callose plugs. *Sexual Plant Reproduction*, 15(4), 195–204.
- Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. and Jørgensen, T.J.D.,** (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics*, 4(7), 873–86.
- Leitner, A., Sturm, M. and Lindner, W.,** (2011) Tools for analyzing the phosphoproteome and other phosphorylated biomolecules: a review. *Analytica Chimica*, 703(1), 19–30.
- Leutwiler, S., Barbara R. H.E, Meyerowitz, E.M.,** (1984) The DNA of *Arabidopsis thaliana*. *Mol Gen Genet*, (194), 15–23.
- Li, H.J., Xue, Y., Jia, D.J., et al.,** (2011) POD1 regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis*. *The Plant Cell*, 23(9), 3288–302.
- Lin, Y., Wang, Y., Zhu, J.K. and Yang, Z.,** (1996) Localization of a Rho GTPase implies a role in tip growth and movement of the generative cell in pollen tubes. *The Plant Cell*, 8(2), 293–303.
- Lopes, M. a and Larkins, B. a,** (1993) Endosperm origin, development, and function. *The Plant Cell*, 5(10), 1383–99.

- Lord, E. M.**, (2003) Adhesion and guidance in compatible pollination. *Journal of Experimental Botany*, 54(380), 47–54.
- Luan, S.**, (2002) Tyrosine phosphorylation in plant cell signaling. *Proceedings of the National Academy of Sciences*, 99(18), 11567–9.
- Malek, B. von, Graaff, E. van der, Schneitz, Kay and Keller, B.**, (2002) The *Arabidopsis* male-sterile mutant *dde2-2* is defective in the *ALLENE OXIDE SYNTHASE* gene encoding one of the key enzymes of the jasmonic acid biosynthesis pathway. *Planta*, 216(1), 187–92.
- Malho, R., Read, N.D., Pais, M.S. and Trewavas, A.J.**, (1994) Role of cytosolic free calcium in the reorientation of pollen tube growth. *The Plant Journal*, 5(3), 331–341.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S.**, (2002) The protein kinase complement of the human genome. *Science*, 298(5600), 1912–34.
- Mayank, P., Grossmann, J., Wuest, S., Boisson-Dernier, A., Roschitzki, B., Nanni, P., Nühse, T. and Grossniklaus, Ueli**, (2012) Characterization of the phosphoproteome of mature *Arabidopsis* pollen. *The Plant Journal*, 1–13.
- Mccormick, S.**, (2004) Control of male gametophyte development. *The Plant Cell*, 16(1998), 142–154.
- McCormick, S.**, (1993) male gametophyte development. *The Plant Cell*, 5(10), 1265–1275.
- Miller, D.D., Ruijter, N.C.A. De, Bisseling, T. and Emons, A. Mie C.**, (1999) The role of actin in root hair morphogenesis: studies with lipochito-oligosaccharide as a growth stimulator and cytochalasin as an actin perturbing drug. *The Plant Journal*, 17(2), 141–154.
- Mithoe, S.C. and Menke, F.L.H.**, (2011) Phosphoproteomics perspective on plant signal transduction and tyrosine phosphorylation. *Phytochemistry*, 72(10), 997–1006.
- Miyazaki, S., Murata, T., Sakurai-Ozato, N., Kubo, M., Demura, T., Fukuda, H. and Hasebe, M.**, (2009) *ANXUR1* and 2, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization. *Current Biology*, 19(15), 1327–31.
- Moore, J.M., Calzada, J.P., Gagliano, W. and Grossniklaus, U**, (1997) Genetic characterization of *hadad*, a mutant disrupting female gametogenesis in *Arabidopsis thaliana*. *Cold Spring Harbor Symposia on Quantitative Biology*, 62, 35–47.

- Márton, M.L., Cordts, S., Broadhvest, J. and Dresselhaus, T.,** (2005) Micropylar pollen tube guidance by *egg apparatus 1* of maize. *Science*, 307(5709), 573–6.
- Nühse, T.S., Stensballe, A., Jensen, O.N. and Peck, S.C.,** (2004) Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *The Plant Cell*, 16(9), 2394–405.
- Okuda, S., Tsutsui, H., Shiina, K., et al.,** (2009) Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61.
- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M.,** (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5), 376–86.
- Ong, S.-E., Mittler, G. and Mann, M.,** (2004) Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. *Nature Methods*, 1(2), 119–26.
- Otto, S.P.,** (2003) The advantages of segregation and the evolution of sex. *Genetics*, 164(3), 1099–118.
- Page, D.R. and Grossniklaus, Ueli,** (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nature Reviews. Genetics*, 3(2), 124–36.
- Paradela, A. and Albar, J.P.,** (2008) Advances in the analysis of protein phosphorylation. *Journal of Proteome Research*, 7(5), 1809–18.
- Peck, S.C.,** (2006) Phosphoproteomics in *Arabidopsis*: moving from empirical to predictive science. *Journal of Experimental Botany*, 57(7), 1523–7.
- Pina, C., Pinto, F., Feijó, J.A. and Becker, J.,** (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology*, 138(June), 744–756.
- Pinkse, M.W.H., Uitto, P.M., Hilhorst, M.J., Ooms, B. and Heck, A.J.R.,** (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Analytical Chemistry*, 76(14), 3935–43.
- Preuss, D.,** (1995) Being fruitful : genetics of reproduction in *Arabidopsis*. , 11(4), 524–530.

**Qin, Y. and Yang, Zhenbiao**, (2011) Rapid tip growth: insights from pollen tubes. *Seminars in Cell & Developmental Biology*, 22(8), 816–24.

**Ravan.P.H., Evert.R.F. and Eichhorn.S.E.**, (1999) *Biology of Plants*, 7<sup>th</sup> Edition.

**Regan, S.M. and Moffatt, B. a.**, (1990) Cytochemical analysis of pollen development in wild-type *Arabidopsis* and a male-sterile mutant. *The Plant Cell*, 2(9), 877–889.

**Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W. and Baginsky, S.**, (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiology*, 150(2), 889–903.

**Ross, P.L., Huang, Y.N., Marchese, J.N., et al.**, (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics*, 3(12), 1154–69.

**Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, Frédéric and Faure, J.-E.**, (2003) Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*. *Current Biology*, 13(5), 432–6.

**Samuel, M. a, Tang, W., Jamshed, M., et al.**, (2011) Proteomic analysis of Brassica stigmatic proteins following the self-incompatibility reaction reveals a role for microtubule dynamics during pollen responses. *Molecular & Cellular Proteomics*, 10(12).

**Sanders, L.C. and Lord, E M**, (1989) Directed movement of latex particles in the gynoecia of three species of flowering plants. *Science*, 243(4898), 1606–8.

**Schmidt, S.R., Schweikart, F. and Andersson, M.E.**, (2007) Current methods for phosphoprotein isolation and enrichment. *Journal of Chromatography*, 849(1-2), 154–62.

**Schopfer, C. R.**, (1999) The male determinant of self-incompatibility in Brassica. *Science*, 286(5445), 1697–1700.

**Schwartz, D. and Gygi, S.P.**, (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature Biotechnology*, 23(11), 1391–8.

**Shiba, H.**, (2001) A pollen coat protein, SP11/SCR, determines the pollen S-specificity in the self-incompatibility of Brassica species. *Plant Physiology*, 125(4), 2095–2103.

**Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K. and Ishihama, Y.,** (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Molecular Systems Biology*, 4, 193.

**TAIR-Genome,** (2000) The *Arabidopsis* Initiative.

**Thingholm, T.E., Jensen, O.N., Robinson, P.J. and Larsen, M.R.,** (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Molecular & Cellular Proteomics*, 7(4), 661–71.

**Wessel, D. and Flügge, U.I.,** (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry*, 138(1), 141–143.

**Wilhelmi, L.K. and Preuss, D.,** (1996) Self-sterility in *Arabidopsis* due to defective pollen tube guidance. *Science*, 274(5292), 1535–7.

**Yadegari, R. and Drews, G.N.,** (2004) Female Gametophyte Development. *The Plant Cell* , 16, 133–142.

**Zinkl, G.M., Zwiebel, B.I., Grier, D.G. and Preuss, D.,** (1999) Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine. *Development*, 126(23), 5431–40.



## Aim of the Thesis

---

A comprehensive description of the phosphorylation dynamics involved in *Arabidopsis* pollination would have applications in the understanding of the signaling processes required for successful reproduction. Since pollen is the main allergen for type I allergy, and more than 500 million people suffer from seasonal asthma or hay fever, the phosphoprotein map of a mature dehydrated pollen grain will enhance our understanding of the biology underlying pollen allergy. Although, *Arabidopsis* is the most suitable plant model system, phosphoproteomics studies on pollen and pollination-specific tissues pose a technical challenge because of the requirement of a large sample with respect to quantity. The aim of this thesis was to characterize the phosphoproteome of pollination-specific tissue of *Arabidopsis*.

To achieve this goal, I first analysed and characterised the phosphoproteome of mature, dehydrated pollen grain, which is presented in detail in Chapter 2. With no previous phosphoproteomics study conducted on *Arabidopsis* pollen, this thesis reports the novel phosphoproteome data generated by the application of three different enrichment techniques. Furthermore, in Chapter 2 I also discuss the gene ontology categories crucial for pollen development, as identified in an over-representation analysis of our phosphoproteome data.

Later, in Chapter 3, I present a label-free quantification map of phosphoproteins/phosphopeptides extracted from *Arabidopsis* pollination-specific tissues (pollen, stigma and pistil). In order to characterize dynamic phosphorylation processes, I performed an analysis of the differential expression and a change in abundance of phosphopeptides/phosphoproteins between the unpollinated pistils and pistils 1 hour after pollination using Progenesis LC-MS.

## Materials and Methods

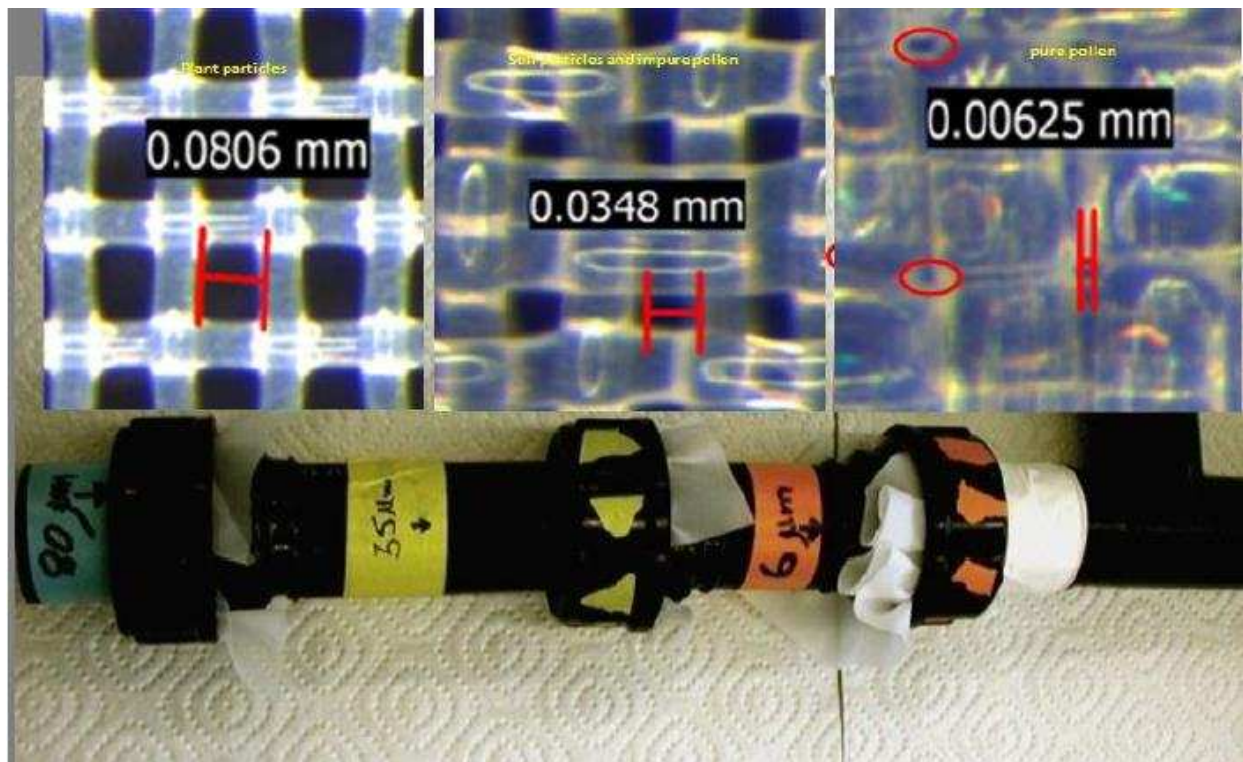
### 3.1 NOTE:

This thesis contains chapters, which individually contain a detailed material and method section. Here I list in table where the various methods are described.

S.No	Materials and Methods	Chapter
1.	Plant material and growth condition	4 & 5
2.	Pollen collection	below
3.	Protein extraction and digestion	4
4.	Pre-fractionation of peptides	4
5.	Phosphopeptides enrichment	4 & 5
6.	Mass-spectrometry	4 & 5
7.	Data analysis tools	4 & 5
8.	Label-free quantification (PROGENESIS LC-MS)	5

### 3.2 Pollen collection

Mature dehydrated *Arabidopsis* pollen grain (Columbia accession) were collected using the “vacuum cleaner” method described previously (Holmes-Davis *et al.*, 2005). An in-house made pipe, equipped with three different meshes of decreasing diameter (80  $\mu\text{m}$  (holds the floral particles), 35  $\mu\text{m}$  (holds the sand particles) and finally 6  $\mu\text{m}$  (holds the pure pollen grains)) was attached to a vacuum cleaner as shown in (Figure 2.1).



**Figure 2.1:** Vacuum cleaner used for pollen collection. A customized pipe fitted with three different meshes of decreasing diameter (80-35-6  $\mu\text{m}$ ) was attached to hand held vacuum cleaner.

With the arrival of first inflorescence, vacuuming was initiated and carried out for 2 weeks every 2<sup>nd</sup> day. In between the plant was allowed to recover from the stress of vacuuming. The collected pollen grains were checked under the microscope to assess their purity and stored at  $-80^{\circ}\text{C}$ .

# Characterization of the phosphoproteome of mature *Arabidopsis* pollen

the plant journal

SEB  
Society for  
Experimental Biology

*The Plant Journal* (2012)

doi: 10.1111/j.1365-3113X.2012.05061.x

## Characterization of the phosphoproteome of mature *Arabidopsis* pollen

Pururawa Mayank<sup>1,2</sup>, Jonas Grossman<sup>3</sup>, Samuel Wuest<sup>4,†</sup>, Aurélien Boisson-Dernier<sup>2</sup>, Bernd Roschitzki<sup>3</sup>, Paolo Nanni<sup>3</sup>, Thomas Nühse<sup>5</sup> and Ueli Grossniklaus<sup>1,2,\*</sup>

<sup>1</sup>Centre for Model Organism Proteomes, University of Zürich, Zürich, Switzerland,

<sup>2</sup>Institute of Plant Biology & Zürich-Basel Plant Science Centre, University of Zürich, Zürich, Switzerland,

<sup>3</sup>Functional Genomics Centre Zürich, Swiss Federal Institute of Technology & University of Zürich, Zürich, Switzerland,

<sup>4</sup>Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, and

<sup>5</sup>Faculty of Life Sciences, University of Manchester, Manchester, UK

Received 19 March 2012; revised 13 May 2012; accepted 21 May 2012.

\*For correspondence (e-mail grossnik@botinst.uzh.ch).

†Present address: Institute of Evolutionary Biology and Environmental Studies & Institute of Plant Biology, Zürich-Basel Plant Science Centre, University of Zürich, Zürich, Switzerland.

## Characterization of the phosphoproteome of mature *Arabidopsis* pollen

Pururawa Mayank<sup>1,2</sup>, Jonas Grossman<sup>3</sup>, Samuel Wuest<sup>4,†</sup>, Aurélien Boisson-Dernier<sup>2</sup>, Bernd Roschitzki<sup>3</sup>, Paolo Nanni<sup>3</sup>, Thomas Nühse<sup>5</sup> and Ueli Grossniklaus<sup>1,2,\*</sup>

<sup>1</sup>Centre for Model Organism Proteomes, University of Zürich, Zürich, Switzerland,

<sup>2</sup>Institute of Plant Biology & Zürich-Basel Plant Science Centre, University of Zürich, Zürich, Switzerland,

<sup>3</sup>Functional Genomics Centre Zürich, Swiss Federal Institute of Technology & University of Zürich, Zürich, Switzerland,

<sup>4</sup>Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, and

<sup>5</sup>Faculty of Life Sciences, University of Manchester, Manchester, UK

Received 19 March 2012; revised 13 May 2012; accepted 21 May 2012.

\*For correspondence (e-mail grossnik@botinst.uzh.ch).

†Present address: Institute of Evolutionary Biology and Environmental Studies & Institute of Plant Biology, Zürich-Basel Plant Science Centre, University of Zürich, Zürich, Switzerland.

### SUMMARY

Successful pollination depends on cell–cell communication and rapid cellular responses. In *Arabidopsis*, the pollen grain lands on a dry stigma, where it hydrates, germinates and grows a pollen tube that delivers the sperm cells to the female gametophyte to effect double fertilization. Various studies have emphasized that a mature, dehydrated pollen grain contains all the transcripts and proteins required for germination and initial pollen tube growth. Therefore, it is important to explore the role of post-translational modifications (here phosphorylation), through which many processes induced by pollination are probably controlled. We report here a phosphoproteomic study conducted on mature *Arabidopsis* pollen grains with the aim of identifying potential targets of phosphorylation. Using three enrichment chromatographies, a broad coverage of pollen phosphoproteins with 962 phosphorylated peptides corresponding to 598 phosphoproteins was obtained. Additionally, 609 confirmed phosphorylation sites were successfully mapped. Two hundred and seven of 240 phosphoproteins that were absent from the PhosPhAt database containing the empirical *Arabidopsis* phosphoproteome showed highly enriched expression in pollen. Gene ontology (GO) enrichment analysis of these 240 phosphoproteins shows an over-representation of GO categories crucial for pollen tube growth, suggesting that phosphorylation regulates later processes of pollen development. Moreover, motif analyses of pollen phosphopeptides showed an over-representation of motifs specific for  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinases, mitogen-activated protein kinases, and binding motifs for 14-3-3 proteins. Lastly, one tyrosine phosphorylation site was identified, validating the TDY dual phosphorylation motif of mitogen-activated protein kinases (MPK8/MPK15). This study provides a solid basis to further explore the role of phosphorylation during pollen development.

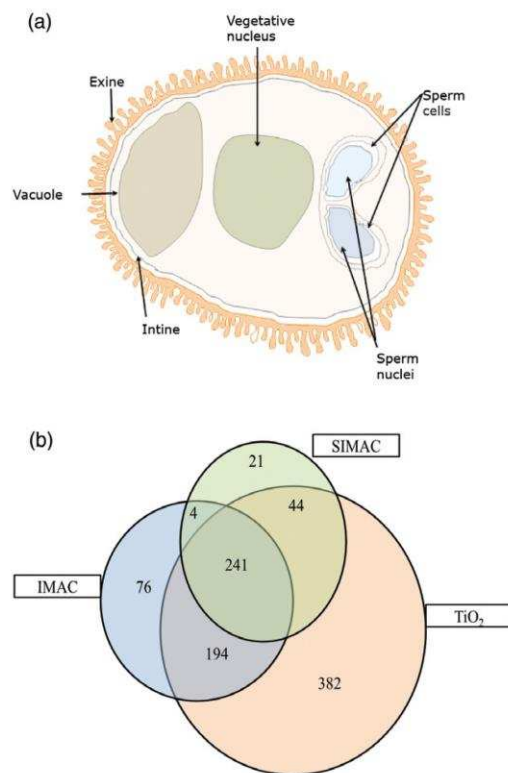
**Keywords:** phosphoproteome, male gametophyte, IMAC, SIMAC,  $\text{TiO}_2$ , kinases.

### INTRODUCTION

The plant lifecycle alternates between two generations, the diploid sporophyte and the haploid gametophytes, which produce the gametes. The mature male gametophyte (pollen grain) represents an independent, highly simplified organism that is specialized to deliver the male gametes to the female gametophyte (embryo sac) (reviewed by Suzuki, 2009; Ma and Sundaresan, 2010). The mature pollen typically consists of a large vegetative cell that harbors two sperm cells in its cytoplasm (Figure 1a). During pollination,

the mature pollen grain is deposited on a stigmatic papillar cell, where it hydrates, germinates, and grows a pollen tube (PT) that transports the two sperm cells to the female gametophyte, ensuring double fertilization. Therefore, even in the dehydrated state, the mature pollen grain must be prepared for the rapid physiological changes occurring during pollination. Genetic studies and experiments using pharmacological inhibitors have shown that transcription and translation are not essential for pollen germination and





**Figure 1.** *Arabidopsis thaliana* male gametophyte phosphoproteome. (a) Schematic of a pollen grain showing the vegetative cell enclosed by the intine and exine, containing a large vacuole and nucleus, and the two sperm cells. (b) Broad coverage was obtained by use of three chromatographic methods. Representation of unique and overlapping phosphopeptides obtained from the soluble protein fraction using the three chromatographic methods.

initial PT growth (Mascarenhas, 1966; Onodera *et al.*, 2008). Indeed, transcriptomics and proteomics studies indicated that a mature pollen grain contains a complex array of transcripts and proteins, allowing germination and rapid growth without *de novo* transcription (Honys and Twell, 2004; Holmes-Davis *et al.*, 2005; Pina *et al.*, 2005; Borges *et al.*, 2008; Wang *et al.*, 2008; Grobei *et al.*, 2009). Thus, the rapid responses during hydration and germination are most likely regulated at the post-translational level, e.g. by phosphorylation.

Phosphorylation is one of the key post-translational modifications that regulate protein functions in diverse biological pathways and contexts. It is a highly dynamic and reversible modification involved in many signaling events and controlled by the activity of a large number of kinases (phosphorylation) and phosphatases (dephos-

phorylation). Phosphorylation can control protein–protein interactions, protein–membrane interactions, protein sub-cellular localization, and protein stability in response to external or internal stimuli (Mithoe and Menke, 2011). As individual phosphoresidues can have distinct functions, it is not sufficient to know whether a protein is phosphorylated or not; rather, it is important to know which particular residue, i.e. serine (Ser), threonine (Thr) or tyrosine (Tyr), is phosphorylated in a phosphopeptide. *In vivo* phosphorylation sites are the direct result of activity of a specific kinase, and, in the case of kinases themselves, the presence of a phosphorylated site (phosphosite) may reflect activation or inactivation of the kinase. Moreover, identifying phosphosites can guide site-directed mutagenesis studies to explore the function of particular proteins. Various means have been proposed to extract phosphorylation motifs from large sets of phosphorylated peptides. Motif-X (Schwartz and Gygi, 2005) extracts significantly over-represented motifs from large-scale phosphoproteomics data, and thus provides valuable information on the kinase–substrate specificity. In eukaryotes, Ser, Thr and Tyr are the most common targets of phosphorylation. Despite the fact that Tyr phosphorylation is less abundant in eukaryotes compared to Ser/Thr phosphorylation, it is crucial for protein–protein interactions (Manning *et al.*, 2002; Dissmeyer and Schnittger, 2011). The human genome encodes for approximately 500 kinases and 100 phosphatases, accounting for almost 2.0% of the genome (Manning *et al.*, 2002). The Arabidopsis genome encodes approximately 1100 kinases and 100–200 phosphatases, accounting for 5% of the genome, which has a similar number of genes as the human genome (Arabidopsis Genome Initiative, 2000). More kinases than phosphatases may be required because protein kinases play a crucial and specific role in the response to a variety of external and internal stimuli, while phosphatases merely reverse the effect of kinases (Smith and Walker, 1996). The vast majority of plant kinases phosphorylate Ser and Thr; only recently have Tyr kinases and their role in signaling events during plant development been uncovered (Luan, 2002; Sugiyama *et al.*, 2008; Jaillais *et al.*, 2011).

In any eukaryotic cell, approximately one-third of the proteins are phosphorylated at any given time (Zolnierowicz and Bollen, 2000; Qeli and Ahrens, 2010). Therefore, large quantities of mature pollen are required to extract enough proteins for phosphopeptide enrichment, as unavoidable sample loss takes place at each enrichment step. However, using optimized enrichment techniques and highly sensitive mass spectrometry, identification of phosphopeptides even from low sample quantities has become feasible. Here, we describe a phosphoproteome atlas of the mature Arabidopsis pollen grain, reporting 962 phosphopeptides corresponding to 598 phosphoproteins. Additionally, the exact position of 609 phosphosites was mapped on these phosphopeptides (Beausoleil *et al.*, 2006), which we thus call

'phosphosite-peptides'. Furthermore, database searches against empirically identified Arabidopsis phosphoproteins (Durek *et al.*, 2010) identified 240 phosphoproteins that were absent from the Arabidopsis phosphorylation database (PhosPhAt). Most of these 240 phosphoproteins are relatively highly expressed in cells of the mature male gametophyte when compared to other tissues from Arabidopsis. A correlative study of the pollen phosphoproteome (this study) with the pollen proteome and transcriptome (Honys and Twell, 2004; Holmes-Davis *et al.*, 2005; Schmid *et al.*, 2005; Borges *et al.*, 2008; Grobei *et al.*, 2009) showed that 96% of the phosphoproteins had previously been identified as *bona fide* pollen proteins, providing additional reassurance that the phosphopeptides were correctly identified.

The dehydrated pollen grain contains all the proteins required for the pollen hydration, germination and initial tube growth that take place once the mature pollen grain lands on the stigma. We assume that these early events require extensive changes in the phosphoproteome, including phosphorylation and dephosphorylation of proteins. The phosphoproteome of mature pollen comprises a large number of phosphorylated proteins that may be subject to changes in phosphorylation upon pollination. In summary, this study enhances our knowledge on the role of phosphorylation in plant reproduction, provides a framework to understand the dynamics of protein phosphorylation during pollination, and suggests processes that may be regulated by phosphorylation during pollen development.

## RESULTS AND DISCUSSION

### Characterization of the male gametophyte phosphoproteome

Phosphoproteome analyses of mature pollen grains require development of a procedure for efficient isolation and broad coverage of phosphoproteins from limited amounts of tissue (see Experimental Procedures, Figure S1). High-accu-

racy mass spectrometry using an LTQ Orbitrap for collision-induced dissociation (CID) and electron transfer dissociation (ETD) in combination with high-throughput phospho-enrichment chromatography, i.e. immobilized metal ion affinity chromatography (IMAC), titanium dioxide (TiO<sub>2</sub>) chromatography, and sequential elution from IMAC (SIMAC, where the flowthrough of IMAC is passed over a TiO<sub>2</sub> column), were used to obtain a broad phosphoproteome coverage of pollen. The results of the various enrichment procedures are shown in Figure 1(b). TiO<sub>2</sub> chromatography was the most efficient in our analysis; however, the use of SIMAC helped to retain singly phosphorylated peptides, which may be missed because IMAC has high selectivity for multiply phosphorylated peptides (Thingholm *et al.*, 2008). Strong cation exchange (SCX), which has selectivity for positively charged peptides, was used for pre-fractionation. Negatively charged phosphopeptides elute in the initial fractions (Figure S2). The enriched sample was loaded onto an LTQ Orbitrap, and the resulting data files from mass spectrometry were analyzed using the MASCOT search algorithm (Perkins *et al.*, 1999). We identified a total of 962 unique phosphopeptides corresponding to 598 unique phosphoproteins with a false discovery rate of 0.3% (spectra level) from soluble and membrane fractions collectively (Table S1). To extract maximum information from our phosphopeptides, we used PeptideClassifier, a generic deterministic peptide classification scheme (Grobei *et al.*, 2009; Qeli and Ahrens, 2010). This classified 94% of the pollen phosphopeptides into high information content classes (classes 1A, 1B, 2A and 2B; see Supporting Information Data S1), which means that a protein sequence and gene locus were unambiguously identified for most of the phosphopeptides (Table 1).

In a phosphopeptide, multiple phosphosites are often seen that may potentially be phosphorylated. Based on the fragment ions, it is impossible to confidently distinguish them, and we therefore used a heuristic approach to

**Table 1** Peptide classifier analysis

Peptide classes	Number of spectra (pre-validation)	Number of spectra (post-validation)	Peptide sequences (pre-validation)	Peptide sequences (post validation)	Number of proteins (pre-validation)	Number of proteins (post-validation)
1A	20 599	20 585	591	577	363	349
1B	3247	3247	79	79	45	45
2A	112	112	8	8	8	8
2B	5187	5168	252	233	168	149
3A	40	40	3	3	2	2
3B	3009	3009	62	62	83-212	83-212
Total	32 194	32 161	995	962	669*	598*

The phosphopeptides were classified into various classes based on their information content (Qeli and Ahrens, 2010; see Supporting Information Data S1). The total number of spectra for each evidence class, the number of distinct peptides and the number of proteins (pre- and post-validation of the single-hit protein identifications) per class are shown. For class 3B, we are unable to give an exact protein count due to ambiguous assignments: the least number of proteins explaining these peptides is 83 proteins, while the maximum number is 212 proteins.

\*The total number of proteins pre- and post-validation including splice variants based on the minimal number of class 3B proteins.

© 2012 The Authors

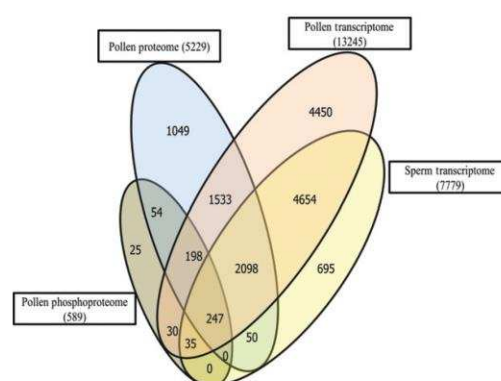
The Plant Journal © 2012 Blackwell Publishing Ltd, *The Plant Journal*, (2012), doi: 10.1111/j.1365-3113X.2012.05061.x



determine whether we could confirm a specific phosphorylation site within a phosphopeptide. (Elias *et al.*, 2004). Phosphorylation sites were mapped on approximately 600 phosphopeptides and were named either 'phosphosite-peptides' (peptides with confirmed phosphorylation sites) or 'phosphopeptides' (if the phosphorylation site could not be unambiguously located). We found 934 singly, 64 doubly and two multiply phosphorylated peptides. The sites of phosphorylation were 86% Ser, 14% Thr and 0.16% Tyr (one site), which is comparable to previous studies (Reiland *et al.*, 2009). Different frequencies of Tyr phosphorylation are seen in various studies (Nühse *et al.*, 2004; de la Fuente van Bentem *et al.*, 2008; Sugiyama *et al.*, 2008; Reiland *et al.*, 2009), and most likely depend on the organisms, tissues and methodologies used to enrich the phosphopeptides. Nevertheless, Ser and Thr phosphorylations are much more frequent in plants when compared to Tyr phosphorylation, as Ser/Thr kinases are commonly encoded in plant genomes, while no *bona fide* Tyr kinases have been found (Sugiyama *et al.*, 2008). Thus, Tyr phosphorylation was thought to be limited to a few known dual-specificity kinases, and, until recently, no functional importance for Tyr (de)phosphorylation in plants had been reported. However, two recent studies showed that Tyr dephosphorylation of BRASSINOSTEROID-INSENSITIVE2 (BIN2), a glycogen synthase kinase 3-like kinase, and Tyr phosphorylation of BRASSINOSTEROID INSENSITIVE1 KINASE INHIBITOR1 (BKI1) are crucial for downstream regulation of brassinosteroid signaling (Kim *et al.*, 2009; Jaillais *et al.*, 2011).

To assess the quality of the pollen phosphoproteome dataset, we compared it with the pollen proteome (Grobei *et al.*, 2009) (downloaded from Pep2Pro, which is based on the latest TAIR10 annotations) and the pollen transcriptome (re-analyzed with TAIR9, as the genome sequence and annotation have evolved significantly after initial design of probe sets for the Affymetrix ATH1 array). The latter includes transcriptomes of early stages of microgametogenesis (Honys and Twell, 2004), sperm cells (Borges *et al.*, 2008), and mature pollen (Schmid *et al.*, 2005; Borges *et al.*, 2008) (Figure 2). Interestingly, this comparison revealed that 96% of the identified phosphoproteins have either corresponding transcripts or proteins detected in previous transcriptomics or proteomics studies, thus validating the quality of our experimental approach.

Furthermore, to broaden our understanding of pollen biology, we compiled a list of genes required for pollen development (Table S3). Mutations in these genes cause a defect in pollen development and/or function. Interestingly, 18 genes (Table S2) of a total of 224 genes, which when mutated showed phenotype during pollen development (Table S2), were found in our phosphoproteome dataset. For 12 of these 18 genes, a total of 15 phosphosite-peptides were identified, which may be involved in regulation by phosphorylation and/or dephosphorylation (Table S2). In



**Figure 2.** Overlap of the pollen phosphoproteome with the transcriptome and proteome.

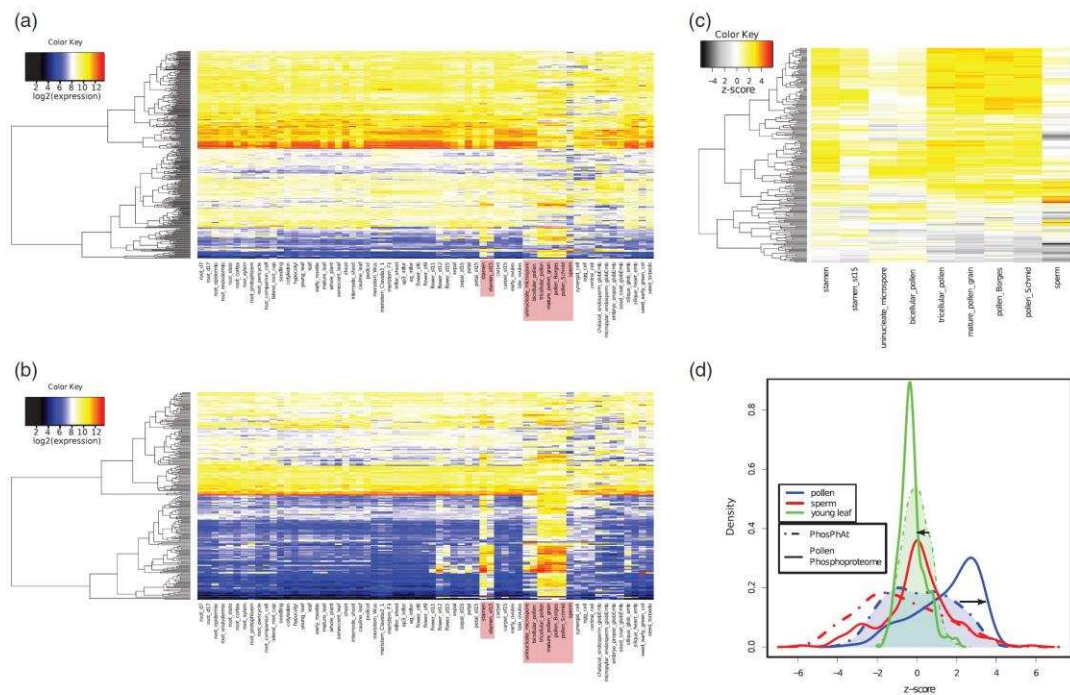
Venn diagram visualizing the overlap of the mature dehydrated Arabidopsis pollen phosphoproteome dataset with previous proteomics (Grobei *et al.*, 2009) and transcriptomics studies conducted on Arabidopsis pollen (Honys and Twell, 2004) and sperm (Borges *et al.*, 2008). Approximately 96% of the phosphoproteins excluding the splice variant have been detected in previous proteomics and transcriptomics studies; for the remaining 4%, there is either no probe on the ATH1 array or they were not found in these large-scale pollen analyses.

order to provide support for a possible regulatory role of these phosphosites, we checked whether the phosphoresidue was conserved in the orthologous proteins of two other monocot species (*Zea mays* and *Oryza sativa*) and two dicot species (*Vitis vinifera* and *Populus trichocarpa*). The amino acid sequence of context six Arabidopsis phosphosites was conserved in all five species (Table S2), indicating a functional role. Thus, such information will be useful in planning future site-directed mutagenesis studies to explore the role of phosphorylation in genes known to be essential for pollen development.

#### A subset of 240 phosphoproteins are enriched in pollen-specific tissues and cells

A comparative study between our pollen phosphoprotein dataset and the PhosPhAt database, which contains experimentally identified Arabidopsis phosphopeptides and phosphoproteins (Heazlewood *et al.*, 2008; Durek *et al.*, 2010), identified a subset of 240 phosphoproteins that were absent from PhosPhAt (as of 17 May 2011; data kindly provided by Waltraud Schulze, Max-Planck-Institut für Molekulare Pflanzenphysiologie, Golm, Germany). As PhosPhAt contains data generated in many large-scale studies, the absence of these 240 phosphoproteins (Table S4), corresponding to 6% of the PhosPhAt dataset, had to be explained. To our knowledge, none of the large-scale studies compiled in PhosPhAt used pollen, probably explaining the absence of these 240 phosphoproteins from PhosPhAt. To confirm this hypothesis, we checked their expression in an expression atlas containing a large number of Arabidopsis





**Figure 3.** Phosphoproteins not found in PhosPhAt are largely pollen-enriched.

(a, b) Representation of relative expression values (per gene-normalized; also termed z-scores) across the tissue atlas (Wuest *et al.*, 2010) for genes encoding phosphoproteins that are (a) present or (b) absent in PhosPhAt, with black/blue indicating low expression and yellow/red indicating high expression. The tissues were sorted according to developmental stage, and genes were clustered based on euclidean distance.

(c) Representation of relative expression values (row-scaled; also termed z-scores) expression values across the tissue atlas (shown in Figure 3b), with a focus on male reproductive cells only. Genes were clustered based on euclidean distance. Overall z-scores are shifted towards high values, indicating predominant expression of the genes in male reproductive cells, especially in mature pollen stages and less so in sperm.

(d) Representation of z-score densities for the 207 phosphoproteins absent from PhosPhAt (solid lines) and 310 phosphoproteins present in PhosPhAt (dashed lines). The phosphoproteins that are absent from PhosPhAt are shifted towards higher values in male gametophytic tissues/cells compared to phosphoproteins present in PhosPhAt, indicating predominant expression of the genes during male gametogenesis.

tissues and cell types (Wuest *et al.*, 2010), including various stages of male gametogenesis (Honys and Twell, 2004). The heat maps in Figure 3 show expression estimates across this Arabidopsis tissue atlas of genes encoding phosphoproteins from our dataset that were present (310/358 are represented on the ATH1 array) or absent in PhosPhAt (207/240 are represented on the ATH1 array). Although genes encoding phosphoproteins present in PhosPhAt were expressed in pollen, they did not show a strong enrichment for male reproductive tissues (Figure 3a,d). In contrast, most of the phosphoproteins absent from PhosPhAt displayed high and often enriched expression during male gametogenesis (Figure 3b). Some genes were already expressed at early stages, but their expression increased towards the mature stages of pollen development (Figure 3c).

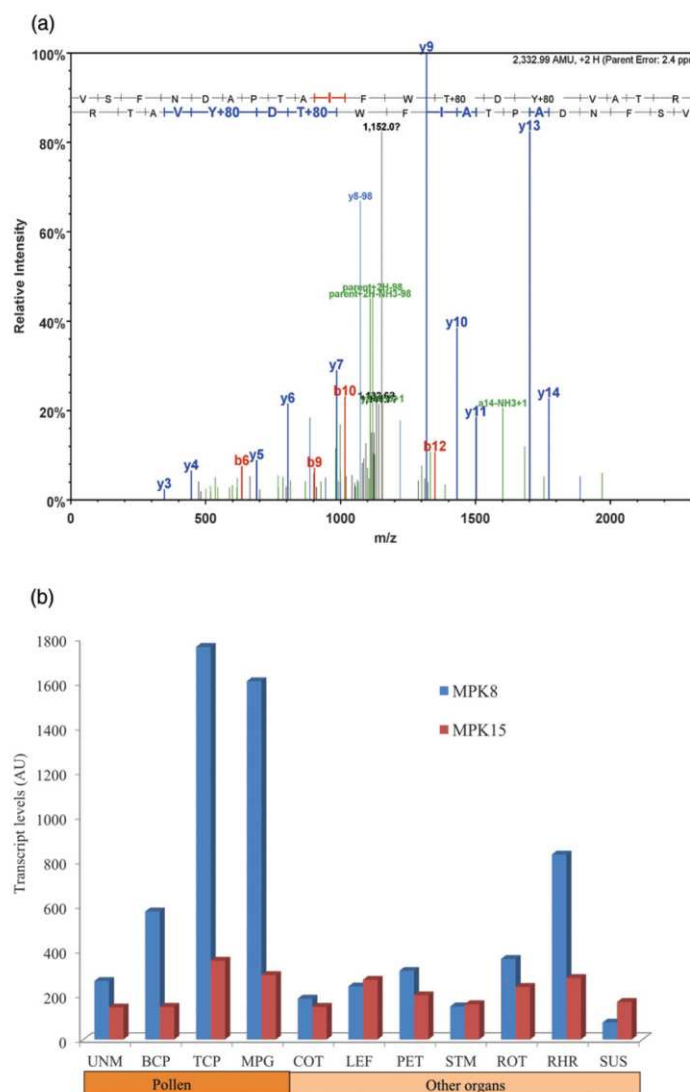
Even though some of the pollen-enriched expression may be due to transcripts predominantly expressed in sperm, a subgroup of transcripts seem strongly depleted from sperm

cells (Figure 3c,d) and are probably only expressed in the vegetative cell, which is not separately represented in the Arabidopsis tissue atlas. Possibly fewer phosphorylation events occur in mature sperm, which is metabolically much less active, if not inactive, and does not interact with sporophytic tissues such as the vegetative cell, which produces the PT. Finally, a small subset of genes showed a high expression in the stamen, which contains the developing microspores (pollen). The corresponding proteins are probably derived from pollen or, alternatively, may have been synthesized in the tapetum, a cell layer of the anther that surrounds the microspores and actively participates in pollen exine formation (Ariizumi and Toriyama, 2011). Alternatively, we cannot totally exclude some minor contamination by sporophytic cells from stamen debris during sample collection. Nonetheless, the 240 phosphoproteins absent from PhosPhAt will hereafter be referred to as pollen-enriched phosphoproteins.

### Dual phosphorylation of MPK8 at the TDY motif in pollen

Among the 962 unique phosphopeptides, we found only one Tyr-phosphorylated peptide [VSFNDAPTAIFWT\*DY\*VATR], which is in PeptideClassifier class 3B. The phosphopeptide could thus not unambiguously be assigned to a gene model,

and corresponds to either mitogen-activated protein kinase (MAPK) MPK8 (AT1G18150) or its close homolog MPK15 (AT1G73670). Tyr phosphorylation was validated by manually assigning fragment ions in the spectra to the expected ion type. Furthermore, Tyr phosphorylation was confirmed by using an additional search engine, X-tandem (Fenyo and



**Figure 4.** Dual phosphorylation of MPK8/MPK15 at the TDY motif in pollen.

(a) Tyrosine phosphorylation was confirmed by using the additional search engine X-tandem (Fenyo and Beavis, 2003) with a  $-\log_{10}$  expectation value of 4.54, showing the TDY motif of MPK8/MPK15.

(b) Expression of MPK8/MPK15 in the tissue atlas of Arabidopsis previously generated by Honys and Twell (2004). Expression of MPK8 is highest in tricaric pollen (TCP) and the mature pollen grain (MPG). However, it is expressed throughout many other Arabidopsis tissues [uninucleate microspore (UNM), bicellular pollen (BCP), cotyledon (COT), leaf (LEF), petals (PET), stem (STM), root (ROT), root hair (RHR), suspension cell cultures (SUS)]. On the other hand, MPK15 shows an almost equal expression pattern across all tissues.

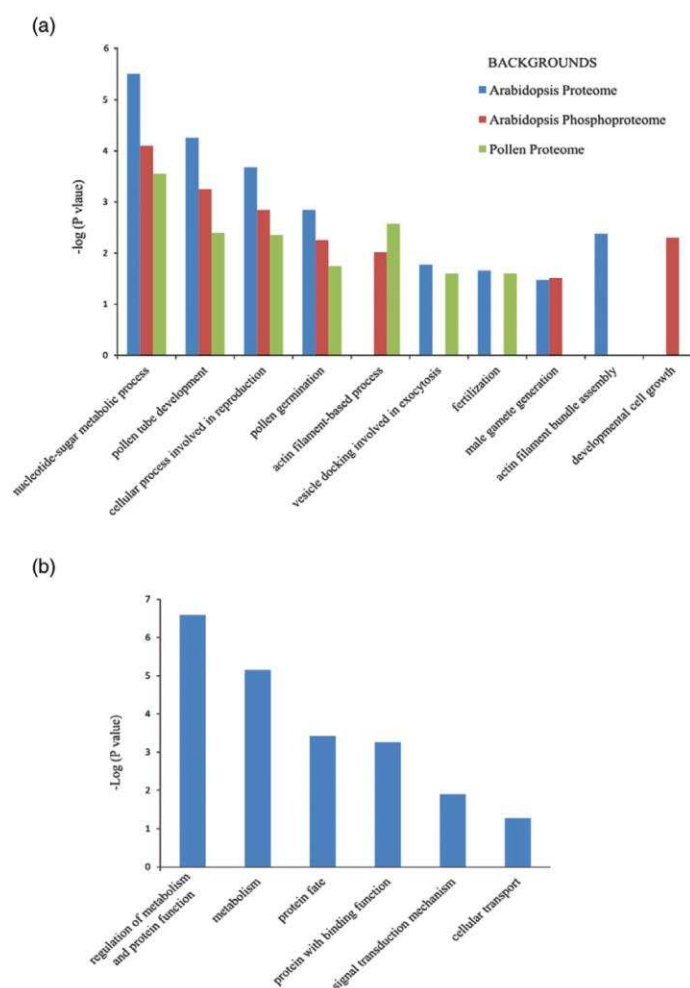


Beavis, 2003), with a  $-\log_{10}$  expectation value of 4.54 (Figure 4a). MAPKs are Ser/Thr-specific protein kinases involved in signaling cascades that respond to extracellular stimuli and regulate various cellular activities, such as development, differentiation, innate immunity, cell death and stress responses (for reviews, see Colcombet and Hirt, 2008; Rodriguez *et al.*, 2010). MAPK cascades are usually initiated by activation of plasma membrane-localized MAPK kinase kinases (MAPKKKs), which activate and phosphorylate MAPKKs, that in turn phosphorylate MAPKs, which then regulate various intracellular effectors (Rodriguez *et al.*, 2010). Around 20 MAPKs are present in the Arabidopsis genome (referred to as MPKs), and they are grouped into four sub-classes (A–D), based on their sequence similarity (Rodriguez *et al.*, 2010). MPK8 and MPK15 belong to sub-group D, defined by a conserved canonic TDY motif instead of the typical TEY motif shared by sub-groups A, B and C (MAPK Group, 2002). This motif is thought to be phosphorylated by MAPKKs. Recently, it was reported that MPK8 activity is increased by mechanical wounding, and that MPK8 regulates the homeostasis of reactive oxygen species in seedlings (Takahashi *et al.*, 2011). Full activity of MPK8 required both calmodulins and MKK3, a MAPKK. Moreover, using a phosphotyrosine antibody, MKK3-dependent phosphorylation of MPK8 at a tyrosine residue has been demonstrated (Takahashi *et al.*, 2011). Although phosphorylation of MAPKs at the Thr and Tyr residues in the TDY motif is reported in other organisms, e.g. mouse, rat and humans (Fu *et al.*, 2005), there was no definitive evidence in plants. Here, we confirm the dual phosphorylation of MPK8 or MPK15 in Arabidopsis pollen at Thr and Tyr residues in the TDY motif (Figure 4a). MPK8 expression is quite high in pollen when compared to other Arabidopsis tissues, while MPK15 expression appears to be low in all tissues examined (Figure 4b). In particular, expression of MPK8 increases from the uninucleate microspore to mature pollen (Honys and Twell, 2004). Although no role for MPK8 or MPK15 in pollen has been described so far, reactive oxygen species have been shown to be involved in PT germination and growth in tobacco and kiwi fruit, and may play a role in pollen–stigma interactions (McInnis *et al.*, 2006; Potocky *et al.*, 2007; Speranza *et al.*, 2011). Further studies are required to test whether MPK8 and/or MPK15 also regulate reactive oxygen species levels in Arabidopsis pollen. Interestingly, in addition to the dual phosphorylation at the TDY motif, we also found a phosphosite that could be unambiguously assigned to MPK15 at Ser511 (TTAAVASTLDS\*PK), a site not previously known to be phosphorylated. Moreover, a phosphopeptide (LASISGDALILDGQT\*PR) for MAP3K $\epsilon$ 2 (MAPKKK6, AT3G07980), which is functionally redundant with MAP3K $\epsilon$ 1 (MAPKKK7, AT3G13530) and known to play a crucial role in pollen viability (Chaiwongsar *et al.*, 2006), was also identified. Our results show that MAPK signaling cascades are active in Arabidopsis pollen and play a role in its development.

#### Over-represented GO and FunCat categories crucial for pollen tip growth

In order to compare the distribution of phosphoproteins in biological processes, we performed a gene ontology (GO) study on the 240 pollen-enriched phosphoproteins using the TopGO algorithm (Alexa *et al.*, 2006). We performed this analysis using the 240 pollen-enriched phosphoproteins as foreground to three distinct backgrounds: the pollen proteome (protein list extracted from Pep2Pro containing the latest TAIR10 annotations; Grobei *et al.*, 2009; Baerenfaller *et al.*, 2011), the total empirical proteome (Pep2Pro, an Arabidopsis tissue-specific proteome database; Baerenfaller *et al.*, 2011), and the total empirical Arabidopsis phosphoproteome (PhosPhAt; Heazlewood *et al.*, 2008; Durek *et al.*, 2010). We looked at the top ten GO categories showing over-representation against one or more backgrounds. Interestingly, this study revealed over-representation of a comparatively large set of gene categories specific to pollen developmental processes, such as ‘pollen germination’, ‘pollen tube development’, ‘cellular process involved in reproduction’, ‘fertilization’, ‘male gamete generation’, ‘actin filament-based processes’, ‘vesicle docking involved in exocytosis’ and ‘metabolic processes’ (Figure 5a). Many of these categories have obvious roles in plant reproduction. Others regulate polarized tip growth of PTs within the female reproductive tract (Hepler *et al.*, 2001). GO categories such as ‘vesicle docking’, ‘monosaccharide metabolic process’ and ‘nucleotide-sugar metabolic processes’ facilitate the cellular processes required to produce enough cell-wall and membrane material for rapid expansion and growth of the PT. Dynamics of actin filaments, involving the categories ‘actin filament based-process’ and ‘actin filament bundle assembly’, is crucial for maintaining the shape of the PT and unidirectional tip growth. Over-representation of these GO categories supports earlier transcriptomics studies (Honys and Twell, 2004; Pina *et al.*, 2005), which showed that physiological activities are at a minimum in mature dehydrated pollen grains, as they mostly contain the transcripts and proteins required for their germination (Honys and Twell, 2004). It was also reported that mature pollen grains express a reduced set of genes for transcription and translation compared to other vegetative tissues (Pina *et al.*, 2005). As most proteins appear to be present in the mature pollen grain, the subsequent developmental processes, such as rehydration, germination, and rapid PT growth, are probably regulated by post-translational events, including phosphorylation and dephosphorylation.

Functional characterization of the pollen-enriched phosphoproteome was achieved using the FunCat algorithm (Surmeli *et al.*, 2008) (Figure 5b). Based on the *P* values, and in agreement with previous transcriptomics (Honys



**Figure 5.** Over-represented GOs and FunCat categories are crucial for pollen tip growth.

(a) Over-representation of GO categories seen in the pollen-enriched phosphoprotein dataset against three different backgrounds: the pollen proteome (protein list extracted from Pep2Pro based on the latest TAIR10 annotations; Grobei *et al.*, 2009; Baerenfaller *et al.*, 2011), the total empirical proteome (Pep2Pro, an Arabidopsis tissue-specific proteome database; Baerenfaller *et al.*, 2011), and the total empirical Arabidopsis phosphoproteome (PhosPhAt; Heazlewood *et al.*, 2008; Durek *et al.*, 2010). The top ten over-represented categories against one or more backgrounds are shown.

(b) Functional analysis of the pollen-enriched phosphoproteins (240) using FunCat (Surmeli *et al.*, 2008) ( $P$  value  $< 0.05$  for over-representation of a FunCat category among pollen-enriched phosphoproteins vs those expected based on the total predicted Arabidopsis proteome).

and Twell, 2004; Pina *et al.*, 2005) and proteomics studies (Grobei *et al.*, 2009), categories involved in metabolism, signal transduction and cellular transport are over-represented in this dataset. This suggests that, even though physiological activities are at a minimum in mature pollen grains, a burst of metabolism is required for germination and initial growth. These early events of PT growth are largely independent of transcription, such that post-translational regulation, involving an extensive network of signaling events, plays a dominant role in these processes.

Our results also show a significant enrichment of categories related to protein fate and protein binding, which are other post-translational events that probably contribute to pollen developmental processes.

#### Phosphorylation of cytoskeleton-associated proteins in pollen

In our study, various classes of proteins with a potential role in PT growth were found to be phosphorylated, providing evidence for post-translational regulation. Proteins associated



with cytoskeletal components, such as actin (VLN5, VLN3 and VLN2) and tubulin (TUA1, TUA3, TUB4 and TUA6) were also phosphorylated. VLN5 and VLN3 are known to be major regulators of actin filament stability and actin bundle formation, and VLN5 has been shown to be required for normal pollen tip growth (Khurana *et al.*, 2010; Zhang *et al.*, 2010). TUA1, which is primarily expressed in pollen, is required for the deposition of cell-wall components during PT growth in *Picea wilsonii* (Yu *et al.*, 2009). However, TUA1 plays no distinct role in tip growth, as it was found to be absent in root hairs that, like PTs, elongate specifically at the tip (Carpenter *et al.*, 1992).

#### Phosphorylation of kinases and receptor-like kinases (RLKs) in pollen

During pollination, when the pollen grain interacts with the papillar cells, rapid recognition, hydration, and germination of the pollen take place. These processes involve cell-cell communication and rapid cellular responses, which are most likely mediated by cytoplasmic kinases and receptor-like kinases (RLKs). In the Arabidopsis genome, more than 1100 kinases are present, of which 620 are annotated as RLKs (Arabidopsis Genome Initiative, 2000; Shiu and Bleecker, 2003). We identified 72 kinases in our list of phosphoproteins (Table S6). Some of the kinases in our list have been shown to play a crucial role in pollen development and PT growth. For example, the AGC kinases AGC1.5 (AT3G12690) and AGC1.7 (AT1G79250) are Ser/Thr kinases that are specifically found at later stages of pollen development and redundantly control polarity in PTs (Zhang *et al.*, 2009). There were also two calcium-dependent protein kinases (CDPKs) in our list: CPK11 (AT1G35670) and CPK4 (AT4G09570), whose role in pollen is not yet elucidated. The following were also identified as targets of phosphorylation in the mature pollen grain: SNF1 kinases (AKIN10, AT3G01090; AKIN11, AT3G29160) and SNF1-related kinases (SnRK2.1, AT5G08590; SnRK2.4, AT1G10940; SnRK2.5, AT5G63650) represent a conserved family of protein kinases known to control metabolic homeostasis, stress signaling, and plant development (Fragoso *et al.*, 2009). Moreover, antisense SnRK1 causes abnormal pollen development in barley, suggesting an important role of this kinase family for pollen function (Zhang *et al.*, 2001). Apart from the TDY motif of MPK8 and/or MPK15 (see above), we also found MAP3K $\epsilon$ 2 to be phosphorylated in our list of mitogen-activated protein kinases. MAP3K $\epsilon$ 2, which is functionally redundant with MAP3K $\epsilon$ 1, is required for proper pollen development and viability (Chaiwongsar *et al.*, 2006).

Most of the known RLKs vary greatly in their domain organization and sequence identity in their extracellular domains, and thus may function in perception of a wide range of stimuli. The Arabidopsis genome encodes more than 400 RLKs, kinases that show a similar domain configuration to trans-membrane receptors, suggesting that they

are involved in cell-surface signal perception (Shiu and Bleecker, 2003). Interestingly, of 46 RLK sub-families, only four were represented in our list, identified by nine phosphopeptides (Table S6). Seven members from three RLK sub-families [RLCKVII (AT5G16500, AT2G07180 and AT3G20530), RLCKIX (AT1G78940, AT1G16760 and AT2G24370) and PERKL (AT3G18810)] were identified by information-rich peptides (PeptideClassifier classes 1A and 2B). The last sub-family, CrRLK1L, was only represented by two class 3B peptides. One of the peptides is perfectly conserved between AT2G21480 and AT4G39110, which are highly expressed in pollen. The second peptide is shared between 11 members of the CrRLK1L sub-family, and we therefore list all these members (Table S6). Some of the CrRLK1L family members, e.g. FERONIA (FER), THESEUS1, ANXUR1 and ANXUR2 (ANX1/2), have been shown to play a role in cell elongation and cell wall-related processes (Boisson-Dernier *et al.*, 2011; Kessler and Grossniklaus, 2011). FER and ANX1/2 appear to act antagonistically in the regulation of PT growth and rupture. In fact, FER non-cell-autonomously inhibits PT growth and promotes rupture of the PT during PT reception, while ANX1/2 act cell-autonomously to redundantly maintain PT growth by inhibiting PT rupture (Escobar-Restrepo *et al.*, 2007; Miyazaki *et al.*, 2009; Boisson-Dernier *et al.*, 2009). As ANX1, ANX2, AT2G21480 and AT4G39110 are the only CrRLK1Ls preferentially expressed in pollen, we allocated the second peptide to them. Interestingly, the two phosphosite-peptides identified in pollen for CrRLK1Ls are the same as those found for CrRLK1Ls in suspension culture cells (Nuhse *et al.*, 2004), and are located in the intracellular kinase domain. Taken together, these results show that phosphorylation of pollen-expressed kinases plays an important role in pollen development, and reflects either kinase auto-phosphorylation or trans-phosphorylation between cooperating kinases.

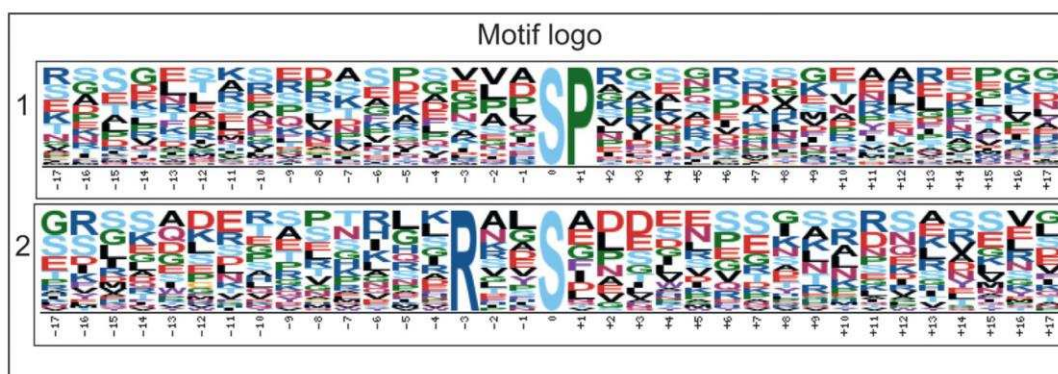
#### Identification of kinase-specific motifs in pollen phosphorylation sites

Phosphorylation sites may be directly responsible for the activation or inactivation of various kinases and other signaling molecules in the cell, and hence can serve as footprints for the various kinase activities. To predict the relationship between the phosphorylation sites and kinase specificity in the pollen phosphoproteome, we performed a motif search using Motif-X. Motif-X is a web-based tool that extracts over-represented patterns of amino acids from the input sequences of phosphopeptides (Schwartz and Gygi, 2005). Each of the over-represented phosphorylation motifs reflects a particular kinase specificity. The over-represented patterns were extracted using the 220 Ser-phosphosite-peptides that were absent from the PhosphoAt database. A motif could not be extracted for Tyr- and Thr-phosphorylated peptides because there were too few of them.

**Table 2** Motif-X analysis

Database and motifs	Foreground matches	Foreground size	Background matches	Background size	Fold increase
PhosPhAt					
(..SP..) motif	322	2468	38682	988763	3.33
(..R..S...) motif	232	1970	43029	942150	2.58
Pollen phosphoproteome					
(..SP..) motif	59	170	51930	992449	6.63
(..R..S...) motif	34	111	51207	940519	5.63

The table above shows the data for over-represented motifs in our datasets (Ser phosphosite-peptides, pollen-enriched phosphoproteome) when compared with PhosPhAt Ser phosphosite-peptides. The Motif-X study conducted on the all Ser phosphosite peptides of PhosPhAt resulted in 21 motifs, XXXXSPXXX (...SP...) and XXXXSXXX (...R..S...) representing the two of them that are over-represented in the pollen-enriched phosphoproteome. These two motifs are found more than twice as often as expected based on PhosPhAt, which was used at background.

**Figure 6.** Identification of kinase-specific motifs in pollen phosphorylation sites using Motif-X.

Alignment of pollen-enriched Ser-phosphorylated peptides. Note that only peptides with confirmed phosphorylation sites were used to acquire the motifs. Motifs 1 and 2 are Motif-X-extracted over-represented motifs from our dataset with a significance value of  $P = 0.000001$  using the IPI Arbidopsis Proteome as background (see Experimental Procedures). Motif 1 is a proline-directed kinase motif that has specificity for MAPKs. Motif 2 is a motif that has specificity for  $\text{Ca}^{2+}$ /CaM-dependent protein kinases and 14-3-3 proteins.

Motif-X extracted two over-represented patterns from the Ser phosphopeptides with a significance value of  $P = 0.000001$  (Table 2, Figure 6). One of the motifs is XXX[pS]PXXX, which suggests proline-directed kinases, such as an MAPKs or cyclin-dependent kinase (Lee *et al.*, 2011). The basophilic motif RXX[pS]XX is a substrate for the  $\text{Ca}^{2+}$ /CaM-dependent protein kinase (CAMK2) family in non-plant organisms (Lee *et al.*, 2011) in the Phospho.ELM database of phosphorylated sites (<http://phospho.elm.eu.org/>), and is often recognized by members of the CDPK-SnRK plant superfamily (Hrabak *et al.*, 2003). A homolog of mammalian CAMK2 is also expressed in plant cells and shows high homology to Ser/Thr receptor-like kinases in plants, which may possibly regulate signal transduction processes. However, not much is known about its role in signal transduction in plants (Watillon *et al.*, 1993; Yang *et al.*, 2004). Previous studies in lily (*Lilium longiflorum*) and tobacco (*Nicotiana tabacum*) show that a chimeric  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase (CCaMK) is expressed in a stage-specific manner in

developing anthers. The expression starts in pollen mother cells and reaches its peak at the tetrad stage, after which a decline in expression was seen; no expression was observed later during pollen development. The stage- and tissue-specific appearance of CCaMK in anthers suggests that it could play a role in changing  $\text{Ca}^{2+}$  concentration in target cells, thereby controlling developmental events in the anther (Poovaiah *et al.*, 1999). The RXX[pS]XX motif is also bound by 14-3-3 proteins, a ubiquitous family of molecules that participate in protein kinase signaling pathways in all eukaryotic cells (Ku *et al.*, 1998) (<http://www.hprd.org/Serine.pdf>). 14-3-3 proteins are phospho-Ser/phospho-Thr binding proteins that form homo- and heterodimers, the biochemical functions of which are to interact with specific target or client proteins, thereby effecting a change in the client. As a result, 14-3-3 proteins are involved in a vast array of processes, such as stress responses, activation of MAPKs, cell-cycle control and apoptosis, and serve as adapters, activators and repressors. In Arabidopsis, specific and efficient reduction



of expression of 14-3-3 isoforms gives rise to a severe developmental delay (Swatek *et al.*, 2011). In plants, the numbers of known 14-3-3 client proteins are increasing drastically, and over 300 have been reported (Denison *et al.*, 2011). The expanding number suggests that 14-3-3 proteins may be involved in many signaling events and physiological processes in plants. In *Arabidopsis* alone, 13 isoforms are present, but these isoforms have a high degree of sequence similarity within or between species (Denison *et al.*, 2011). Ten of the 13 members were present in the pollen proteome (Grobei *et al.*, 2009) and three (GRF1, AT4G09000; GRF2, AT1G78300; GRF12, AT1G78220) were found in the pollen phosphoproteome. This suggests that 14-3-3 proteins and kinases cooperate in a complex signaling network that is crucial for pollen development.

## CONCLUSION

This study on *Arabidopsis* mature pollen grains will help to better understand the signaling processes occurring during pollen development, in particular during pollen hydration, germination and initial PT growth. Some of the phosphoproteins found in our study are strongly enriched in pollen, based on genome-wide expression profiles of the *Arabidopsis* tissue atlas. These proteins can be further explored for their potential role in pollination and PT growth. The data generated in our study also serve as an important reference for downstream signaling events involved in pollination of self-compatible as well as self-incompatible species.

## EXPERIMENTAL PROCEDURES

### Plant material and sample collection

The study was performed using *Arabidopsis thaliana* (L.) Heyhn (accession Columbia). Two collections were performed to generate biological replicates. Plants were grown on a large scale (approximately 10 000) in the greenhouse (22°C during the day and 16°C during the night, 16 h light/8 h dark, humidity 60%) for 4 weeks. With the onset of the first flower, collection of mature pollen was started using a hand-held vacuum-cleaner as described previously (Grobei *et al.*, 2009). Collection of pollen was performed every other day over 2 weeks. After collection, the dry weight of pollen was measured (approximately 30–35 mg), and it was stored at –80°C for further analysis.

### Protein extraction

Protein extraction was performed in two steps. Firstly, tissue was homogenized in liquid nitrogen using a Potter homogeniser (Fisher Scientific, <http://www.fishersci.com/>). Soluble protein was extracted using 100 mM HEPES/KOH pH 7.5, 15 mM EDTA, 25 mM NaF, 0.5% polyvinylpyrrolidone (PVP), 50 mM sodium pyrophosphate (NaPP) and 5% glycerol, plus an inhibitor cocktail for phosphatases (calyculin A, K252a) and proteases (Roche; <http://www.roche.com/>). Secondly, membrane proteins were extracted using 100 and 50 mM triethylammonium bicarbonate (TEAB). All extractions were performed using an Airfuge® (Beckman Coulter, <https://www.beckmancoulter.com>), and the runs were performed at 117–124 kPa for 45 min. Soluble proteins were precipitated by adding four volumes of ice-cold acetone and incubating for 2 h at –20°C. Membrane

proteins were precipitated using methanol/chloroform extraction (Wessel and Flugge, 1984). Protein pellets were briefly dried at 37°C and resuspended in 50 mM TEAB buffer.

### Phosphopeptide enrichment

**Strong cation exchange.** Pre-fractionation was performed by strong cation exchange (SCX). Peptides were desalted using Sep-Pak reverse-phase cartridges (Waters, <http://www.waters.com/>), dissolved in buffer A (7 mM  $\text{KH}_2\text{PO}_4$ , pH 2.65, in 30% acetonitrile), and loaded onto a  $2.1 \times 200$  mm polysulfoethyl aspartamide A column (PolyLC, <http://www.polylc.com/>) on an Agilent HP1100 binary HPLC system (Agilent Technologies, <http://www.home.agilent.com/>). Peptides were eluted in fractions with an increasing KCl gradient (10–40 min, 0–30% buffer B; 40–60 min 30–100% buffer B; buffer B consisted of 7 mM  $\text{KH}_2\text{PO}_4$ , pH 2.65, and 350 mM KCl in 30% acetonitrile). Eluted peptides were pooled into six fractions based on the chromatogram, and desalted using Sep-Pak reverse-phase cartridges. Each fraction was divided into half, and immobilized metal affinity chromatography (IMAC) and titanium dioxide ( $\text{TiO}_2$ ) chromatography were performed simultaneously.

**IMAC.** IMAC was performed as described by Nühse *et al.* (2004) with modifications. Phos-Select resins (Sigma, <http://www.sigma-aldrich.com/>) were used. Resin (40–60  $\mu\text{l}$ ; 50% slurry) was added to a spin column, which was washed with 500  $\mu\text{l}$  equilibration buffer (250 mM acetic acid in 30% acetonitrile). Peptides pre-equilibrated in equilibration buffer were added to the beads and incubated for 1–2 h at room temperature with end-over-end rotation. Peptides were washed twice (15–20 mins each) with equilibration buffer and once with Milli-Q water (Millipore, <http://www.millipore.com/>). Finally, the phosphopeptides were eluted using 200  $\mu\text{l}$  of elution buffer (400 mM ammonium hydroxide).

**$\text{TiO}_2$ .** Chromatography using  $\text{TiO}_2$  (GL Science, <http://www.glscienc.com/>) was performed as described previously (Bodenmiller *et al.*, 2007), with minor modifications. Beads (0.3 mg) were washed with Milli-Q water and 100% methanol. Washed beads were suspended in equilibration buffer (80% acetonitrile + 2.5% trifluoroacetic acid, saturated with phthalic acid). Peptide mixture pre-equilibrated in equilibration buffer was incubated with beads in closed spin columns (MoBiTec, <http://www.mobitec.de/>) for 30 min. After washing twice with saturated phthalic acid solution, twice with 80% acetonitrile + 0.1% trifluoroacetic acid, and finally twice with 0.1% trifluoroacetic acid, the peptides were eluted with 0.3 M  $\text{NH}_4\text{OH}$  (pH 9). Peptides were dried in a speed-vac (Thermoscientific, <http://www.thermoscientific.com/>) and desalted using Zip-Tips ( $\mu\text{C18}$ ; Millipore, <http://www.millipore.com/>) prior to mass spectrometric analysis.

**SIMAC.** Sequential elution from IMAC (SIMAC) was performed on the flowthrough from IMAC as described previously (Thingholm *et al.*, 2008) with minor modifications. The flowthrough was dried in a speed-vac and dissolved in  $\text{TiO}_2$  equilibration buffer.  $\text{TiO}_2$  chromatography was then performed.

### Data acquisition and peptide identification

Dried peptides were desalted and resuspended in 3% acetonitrile and 0.1% formic acid, and analyzed on an LTQ Orbitrap XL mass spectrometer (Thermo Scientific, <http://www.thermoscientific.com>). mgf files were generated from the raw files for all the MS/MS data using Mascot Distiller 2.3 (Berndt *et al.*, 1999) and were selected against a reversed *Arabidopsis* TAIR10 database containing various contaminants (e.g. human keratins and bovine proteins)



using Mascot 2.3. The files were then uploaded to B-fabric (PostgreSQL, <http://www.postgresql.org/>), and finally sorted and manually validated using Microsoft Excel and PivotTable (Microsoft, <http://www.microsoft.com>). For details, see Data S1.

## ACKNOWLEDGEMENTS

We are grateful to Claudia Fortes (Functional Genomics Centre Zürich, Zürich), Valeria Gagliardini, Peter Kopf, Daniela Guthörl, Arturo Bolanos and Karl Huwiler for technical assistance and plant care, Waltraud Schulze (Max Planck Institute of Molecular Plant Physiology, Golm, Germany) for information on PhosPhAt, Katja Bärenfaller (Institut of Agricultural Sciences, Swiss Federal Institute of Technology, Zürich) for providing Pep2Pro data, as well as A. Monica Schauer and Stephen Schauer for helpful suggestions. We thank Ralph Schlappbach (Functional Genomics Centre Zürich, Zürich) for access to the facilities of the Functional Genomics Center Zürich. This work was supported by the University of Zürich, the European Union, and the University Research Priority Program in Functional Genomics/Systems Biology.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Experimental procedures: additional detailed descriptions of protein, mass spectrometric and bioinformatics analyses.

**Figure S1.** Outline of the experimental procedure.

**Figure S2.** Phosphopeptide enrichment profiling.

**Table S1.** Overview of pollen phosphoproteome.

**Table S2.** List of phosphoproteins required for normal pollen development.

**Table S3.** List of published *Arabidopsis* pollen mutants (15 September 2011).

**Table S4.** Pollen-enriched phosphoproteins.

**Table S5.** List of kinases found in the pollen phosphoproteome.

**Table S6.** List of RLKs found in the pollen phosphoproteome.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## REFERENCES

- Alexa, A., Rahnenführer, J. and Lengauer, T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600–1607.
- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Ariizumi, T. and Toriyama, K. (2011) Genetic regulation of sporopollenin synthesis and pollen exine development. *Annu. Rev. Plant Biol.* **62**, 437–460.
- Baerenfaller, K., Hirsch-Hoffmann, M., Svozil, J., Hull, R., Russenberger, D., Bischof, S., Lu, Q., Gruissem, W. and Baginsky, S. (2011) Pep2Pro: a new tool for comprehensive proteome data analysis to reveal information about organ-specific proteomes in *Arabidopsis thaliana*. *Integr. Biol.* **3**, 225–237.
- Beausoleil, S.A., Villen, J., Gerber, S.A., Rush, J. and Gygi, S.P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292.
- Berndt, P., Hobohm, U. and Langen, H. (1999) Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis*, **20**, 3521–3526.
- Bodenmiller, B., Mueller, L.N., Mueller, M., Domon, B. and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods*, **4**, 231–237.
- Boisson-Dernier, A., Kessler, S.A. and Grossniklaus, U. (2011) The walls have ears: the role of plant CrRLK1s in sensing and transducing extracellular signals. *J. Exp. Bot.* **62**, 1581–1591.
- Boisson-Dernier, A., Roy, S., Kritsas, K., Grobei, M.A., Jacubek, M., Schroeder, J.I. and Grossniklaus, U. (2009) Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge. *Development*, **136**, 3279–3288.
- Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J.A. and Becker, J.D. (2008) Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiol.* **148**, 1168–1181.
- Carpenter, J.L., Ploense, S.E., Snustad, D.P. and Silflow, C.D. (1992) Preferential expression of an  $\alpha$ -tubulin gene of *Arabidopsis* in pollen. *Plant Cell*, **4**, 557–571.
- Chaiwongsar, S., Otegui, M.S., Jester, P.J., Monson, S.S. and Krysan, P.J. (2006) The protein kinase genes *MAP3K1* and *MAP3K2* are required for pollen viability in *Arabidopsis thaliana*. *Plant J.* **48**, 193–205.
- Colcombet, J. and Hirt, H. (2008) *Arabidopsis* MAPKs: a complex signalling network involved in multiple biological processes. *Biochem J.* **413**, 217–226.
- Denison, F.C., Paul, A.L., Zupanska, A.K. and Ferl, R.J. (2011) 14-3-3 proteins in plant physiology. *Semin. Cell Dev. Biol.* **22**, 720–727.
- Dissmeyer, N. and Schnittger, A. (2011) The age of protein kinases. *Methods Mol. Biol.* **779**, 7–52.
- Durek, P., Schmidt, R., Heazlewood, J.L., Jones, A., MacLean, D., Nagel, A., Kersten, B. and Schulze, W.X. (2010) PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res.* **38**, D829–D834.
- Elias, J.E., Gibbons, F.D., King, O.D., Roth, F.P. and Gygi, S.P. (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat. Biotechnol.* **22**, 214–219.
- Escobar-Restrepo, J.M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W.C. and Grossniklaus, U. (2007) The *FERONIA* receptor-like kinase mediates male–female interactions during pollen tube reception. *Science*, **317**, 656–660.
- Fenyo, D. and Beavis, R.C. (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **75**, 768–774.
- Fragoso, S., Espindola, L., Paez-Valencia, J., Gamboa, A., Camacho, Y., Martinez-Barajas, E. and Coello, P. (2009) SnRK1 isoforms AKIN10 and AKIN11 are differentially regulated in *Arabidopsis* plants under phosphate starvation. *Plant Physiol.* **149**, 1906–1916.
- Fu, Z., Schroeder, M.J., Shabanowitz, J., Kaldis, P., Togawa, K., Rustgi, A.K., Hunt, D.F. and Sturgill, T.W. (2005) Activation of a nuclear Cdc2-related kinase within a mitogen-activated protein kinase-like TDY motif by auto-phosphorylation and cyclin-dependent protein kinase-activating kinase. *Mol. Cell. Biol.* **25**, 6047–6064.
- de la Fuente van Bentem, S., Anrather, D., Dohnal, I. et al. (2008) Site-specific phosphorylation profiling of *Arabidopsis* proteins by mass spectrometry and peptide chip analysis. *J. Proteome Res.* **7**, 2458–2470.
- Grobei, M.A., Qeli, E., Brunner, E., Rehrauer, H., Zhang, R., Roschitzki, B., Basler, K., Ahrens, C.H. and Grossniklaus, U. (2009) Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Res.* **19**, 1786–1800.
- Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D. and Schulze, W.X. (2008) PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res.* **36**, D1015–D1021.
- Hepler, P.K., Vidali, L. and Cheung, A.Y. (2001) Polarized cell growth in higher plants. *Annu. Rev. Cell Dev. Biol.* **17**, 159–187.
- Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J. and McCormick, S. (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics*, **5**, 4864–4884.
- Hony, D. and Twell, D. (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**, 640–652.
- Hony, D. and Twell, D. (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* **5**, R85.
- Hrabak, E.M., Chan, C.W., Gribskov, M. et al. (2003) The *Arabidopsis* CDPK–SnRK superfamily of protein kinases. *Plant Physiol.* **132**, 666–680.
- Jaillais, Y., Hothorn, M., Belkadir, Y., Dabi, T., Nimchuk, Z.L., Meyerowitz, E.M. and Chory, J. (2011) Tyrosine phosphorylation controls brassinosteroid



- teroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes Dev.* **25**, 232–237.
- Kessler, S.A. and Grossniklaus, U. (2011) She's the boss: signaling in pollen tube reception. *Curr. Opin. Plant Biol.* **14**, 622–627.
- Khurana, P., Henty, J.L., Huang, S., Staiger, A.M., Blanchoin, L. and Staiger, C.J. (2010) *Arabidopsis* VILLIN1 and VILLIN3 have overlapping and distinct activities in actin bundle formation and turnover. *Plant Cell*, **22**, 2727–2748.
- Kim, T.W., Guan, S., Sun, Y., Deng, Z., Tang, W., Shang, J.X., Sun, Y., Burlingame, A.L. and Wang, Z.Y. (2009) Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat. Cell Biol.* **11**, 1254–1260.
- Ku, N.O., Liao, J. and Omary, M.B. (1998) Phosphorylation of human keratin 18 serine 33 regulates binding to 14-3-3 proteins. *EMBO J.* **17**, 1892–1906.
- Lee, T.Y., Lin, Z.Q., Hsieh, S.J., Bretana, N.A. and Lu, C.T. (2011) Exploiting maximal dependence decomposition to identify conserved motifs from a group of aligned signal sequences. *Bioinformatics*, **27**, 1780–1787.
- Luan, S. (2002) Tyrosine phosphorylation in plant cell signaling. *Proc. Natl Acad. Sci. USA*, **99**, 11567–11569.
- Ma, H. and Sundaresan, V. (2010) Development of flowering plant gametophyte. *Curr. Top. Dev. Biol.* **91**, 379–412.
- MAPK Group (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* **7**, 301–308.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
- Mascarenhas, J.P. (1966) Pollen tube growth and RNA synthesis by tube and generative nuclei of *Tradescantia*. *Am. J. Bot.* **53**, 563–569.
- McInnis, S.M., Desikan, R., Hancock, J.T. and Hiscock, S.J. (2006) Production of reactive oxygen species and reactive nitrogen species by angiosperm stigmas and pollen: potential signalling crosstalk? *New Phytol.* **172**, 221–228.
- Mithoe, S.C. and Menke, F.L. (2011) Phosphoproteomics perspective on plant signal transduction and tyrosine phosphorylation. *Phytochemistry*, **72**, 997–1006.
- Miyazaki, S., Murata, T., Sakurai-Ozato, N., Kubo, M., Demura, T., Fukuda, H. and Hasebe, M. (2009) *ANXUR1* and 2, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization. *Curr. Biol.* **19**, 1327–1331.
- Nühse, T.S., Stensballe, A., Jensen, O.N. and Peck, S.C. (2004) Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell*, **16**, 2394–2405.
- Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y. and Pikaard, C.S. (2008) Sex-biased lethality or transmission of defective transcription machinery in *Arabidopsis*. *Genetics*, **180**, 207–218.
- Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.
- Pina, C., Pinto, F., Feijo, J.A. and Becker, J.D. (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.* **138**, 744–756.
- Poovaliah, B.W., Xia, M., Liu, Z., Wang, W., Yang, T., Sathyanarayanan, P.V. and Franceschi, V.R. (1999) Developmental regulation of the gene for chimeric calcium/calmodulin-dependent protein kinase in anthers. *Planta*, **209**, 161–171.
- Potocky, M., Jones, M.A., Bezvoda, R., Smirnov, N. and Zarsky, V. (2007) Reactive oxygen species produced by NADPH oxidase are involved in pollen tube growth. *New Phytol.* **174**, 742–751.
- Qeli, E. and Ahrens, C.H. (2010) PeptideClassifier for protein inference and targeted quantitative proteomics. *Nat. Biotechnol.* **28**, 647–650.
- Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W. and Baginsky, S. (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* **150**, 889–903.
- Rodriguez, M.C., Petersen, M. and Mundy, J. (2010) Mitogen-activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* **61**, 621–649.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**, 501–506.
- Schwartz, D. and Gygi, S.P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398.
- Shiu, S.H. and Bleeker, A.B. (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* **132**, 530–543.
- Smith, R.D. and Walker, J.C. (1996) Plant protein phosphatases. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 101–125.
- Speranza, A., Crinelli, R., Scocciati, V. and Geitmann, A. (2011) Reactive oxygen species are involved in pollen tube initiation in kiwifruit. *Plant Biol.* **14**, 64–76.
- Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K. and Ishihama, Y. (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Mol. Syst. Biol.* **4**, 193.
- Surmeli, D., Ratmann, O., Mewes, H.W. and Tetko, I.V. (2008) FunCat functional inference with belief propagation and feature integration. *Comput. Biol. Chem.* **32**, 375–377.
- Suzuki, G. (2009) Recent progress in plant reproduction research: the story of the male gametophyte through to successful fertilization. *Plant Cell Physiol.* **50**, 1857–1864.
- Swatek, K.N., Graham, K., Agrawal, G.K. and Thelen, J.J. (2011) The 14-3-3 isoforms chi and epsilon differentially bind client proteins from developing *Arabidopsis* seed. *J. Proteome Res.* **10**, 4076–4087.
- Takahashi, F., Mizoguchi, T., Yoshida, R., Ichimura, K. and Shinozaki, K. (2011) Calmodulin-dependent activation of MAP kinase for ROS homeostasis in *Arabidopsis*. *Mol. Cell*, **41**, 649–660.
- Thingholm, T.E., Jensen, O.N., Robinson, P.J. and Larsen, M.R. (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteomics*, **7**, 661–671.
- Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z. and Wu, W.H. (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* **148**, 1201–1211.
- Watillon, B., Kettmann, R., Boxus, P. and Burny, A. (1993) A calcium/calmodulin-binding serine/threonine protein kinase homologous to the mammalian type II calcium/calmodulin-dependent protein kinase is expressed in plant cells. *Plant Physiol.* **101**, 1381–1384.
- Wessel, D. and Flugge, U.I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143.
- Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., von Mering, C. and Grossniklaus, U. (2010) *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* **20**, 506–512.
- Yang, T., Chaudhuri, S., Yang, L., Chen, Y. and Poovaliah, B.W. (2004) Calcium/calmodulin up-regulates a cytoplasmic receptor-like kinase in plants. *J. Biol. Chem.* **279**, 42552–42559.
- Yu, Y., Li, Y., Li, L., Lin, J., Zheng, C. and Zhang, L. (2009) Overexpression of *PwTUA1*, a pollen-specific tubulin gene, increases pollen tube elongation by altering the distribution of  $\alpha$ -tubulin and promoting vesicle transport. *J. Exp. Bot.* **60**, 2737–2749.
- Zhang, Y., Shewry, P.R., Jones, H., Barcelo, P., Lazzeri, P.A. and Halford, N.G. (2001) Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. *Plant J.* **28**, 431–441.
- Zhang, Y., He, J. and McCormick, S. (2009) Two *Arabidopsis* AGC kinases are critical for the polarized growth of pollen tubes. *Plant J.* **58**, 474–484.
- Zhang, H., Qu, X., Bao, C. et al. (2010) *Arabidopsis* VILLIN5, an actin filament bundling and severing protein, is necessary for normal pollen tube growth. *Plant Cell*, **22**, 2749–2767.
- Zolnierowicz, S. and Bollen, M. (2000) EMBO Conference Report. Protein phosphorylation and protein phosphatases. De Panne, Belgium, September 19–24, 1999. *EMBO J.* **19**, 483–488.

## 4.1 Supporting Information

### 4.1.1 Supplemental Experimental Procedures

Additional detailed descriptions of protein, mass spectrometric and bioinformatics analyses.

#### 4.1.1.1 Chemicals

All the chemicals for this study, unless otherwise stated, were obtained from Sigma-Aldrich (<http://www.sigmaaldrich.com/>).

#### 4.1.1.2 In-solution trypsin digestion

1 mg of soluble proteins and 500µg of membrane proteins were used for trypsin digestion. Prior to digestion, samples were reduced with 10 mM DTT for 45 minutes at 55°C and alkylated with 25 mM iodoacetamide for 30 minutes at room temperature in the dark. The reaction was stopped adding 25 mM DTT and incubated at room temperature for 10 minutes. Trypsin (sequencing grade; Promega, <http://www.promega.com/>) was added at a ratio of 1:20 and incubated overnight at 37°C. The reaction was stopped using trifluoroacetic acid (TFA) to a final concentration of 0.5% at 37°C for 30 min followed by centrifugation (13,000g, 10 min). The supernatant was collected for phosphopeptide enrichment.

#### 4.1.1.3 MS analysis and data acquisition

Peptides were desalted and concentrated using ZipTips (µC18; Millipore, <http://www.millipore.com/>). Desalted peptides were vacuum concentrated and afterwards were resuspended in 3% ACN and 0.1% formic acid, and were analyzed on LTQ Orbitrap XL mass spectrometer (Thermo Fischer Scientific, <http://www.thermofisher.com/>) equipped to an Eksigent nano LC system (Eksigent Technologies, <http://www.eksigent.com/>). Peptide separation was made using self-packed (75 µm × 80 mm) reverse phase column packed with C18 material (AQ 3 µM, 200A<sup>0</sup>; Bischoff GmbH, <http://www.bischoff-wohnen.de/>). Solvent composition at the two channels was 0.2% formic acid, 1% ACN for channel A and 0.2% formic acid, 80% ACN for channel B. Peptides were loaded from a Spark Holland auto sampler (<http://www.sparkholland.com/>) and separated using a gradient elution from 3% to 35% ACN in 96 min with a flow rate of 200 nL min<sup>-1</sup>. Full-scan MS spectra (*m/z* 300–2000) were acquired with a resolution of 60000 at *m/z* 400 after accumulation to a target value of 500,000. Up to six data dependent collision induced dissociations (CID) MS/MS spectra were recorded in an ion trap above the threshold of 500, using a normalized collision energy of 35% and an activation time of 30 ms. Charge state screening was enabled and singly charge states were rejected. Precursor masses already selected for MS/MS

were excluded for further selection for 120 s and the exclusion window was set to 20 ppm. The size of the exclusion list was set to a maximum of 500 entries. The mass spectrometer was calibrated using the manufacturer's protocol and during the analysis the orbitrap was recalibrated using internal lock mass on  $m/z$  429.088735 and  $m/z$  445.120025.

#### 4.1.1.4 Data refinement and bioinformatics

mgf files were generated from the raw files for all the MS/MS data with the Mascot distiller 2.3 (Berndt *et al.*, 1999) and were searched against a reversed *Arabidopsis* TAIR 10 database containing contaminants (human keratins and bovine proteins) using MASCOT 2.3 (Matrix Science, <http://www.matrixscience.com/>) (Perkins *et al.*, 1999). Carbamidomethylation (Cys) modification was kept fixed and oxidation (Met) along with carbamidomethyl (Asp, His, Lys, Glu) (N-term), deamidated (NQ), Gln → pyro-Glu (N-term Q), and phosphorylation of Ser, Thr and Tyr was kept as variable modification. A maximum of one missed cleavage on tryptic peptides were allowed. All other precursor ions subjected to MS/MS were searched with a parent mass error tolerance of 10 ppm and a fragment ion error tolerance of 0.8 Dalton. Resulting Mascot result files were parsed into a relational PostgreSQL database of B-fabric (local LIMS) (Turker *et al.*, 2011) using the MS parser library from Matrix Science (<http://www.matrixscience.com/msparser.html>). Peptides scoring with Mascot ion score  $\geq 35$  were accepted for identifications if they were identified on rank one and a significant top ranking (bold red). Then, to increase the quality of the data and to decrease the false discovery rate (FDR), we manually validated the entire single hit MS/MS spectra (168). All protein identifications by peptide assignment were validated manually by two mass spectrometrists and conflicting judgments were validated by the third mass spectrometrists. If the peptide was conserved among more than one protein, then the protein with additional peptides was selected. However, if no additional peptides were identified for the proteins then only one protein was selected (based on *Arabidopsis* Genome Initiative (AGI) ranking).

Furthermore, PeptideClassifier (Grobei *et al.*, 2009; Qeli and Ahrens, 2010) was used to classify the peptides into six different classes (1A, 1B, 2A, 2B, 3A, 3B) based on the information content for each peptide. Class 1A peptides are the most informative as a peptide unambiguously identifies a unique protein sequence and a gene locus. Class 1B peptides unambiguously identify a protein sequence but it can be encoded by several splice isoforms of the same gene model. However, all isoforms of class 1B encode the identical protein sequence, i.e. their transcripts only differ in the 5' and/or 3' UTRs. Class 2 peptides can unambiguously be assigned to a specific gene model but cannot distinguish between several splice isoforms. Class 3A peptides unambiguously identify one protein sequence but this sequence can be encoded by several gene models from distinct loci.

Class 3B peptides cannot be assigned to a particular protein sequence or gene locus and thus represent the least informative class. Therefore, particularly for class 3B peptides, we reported the minimum and maximum number of proteins identified by each class 3B peptide.

Phosphorylation sites mapping on phosphopeptides were confirmed using a normalized delta ion score ( $\Delta I$ ) by taking the difference of Mascot ion score between the rank 1 and rank 2 assignments and dividing that difference by the Mascot ion score of rank 1 phosphopeptide. The phosphorylation site was confirmed when the  $\Delta I$  score  $\geq 0.4$  (Reiland *et al.*, 2009; Elias *et al.*, 2004; Beausoleil *et al.*, 2006)). Phosphopeptides with a confirmed phosphorylation site were named “Phosphosite-peptides”, “PhosphoPEPTIDES” refers to unambiguous peptides for which the phosphorylation site was ambiguous.

#### 4.1.1.5 Microsoft Excel analyses, gene ontology, PhosPhAt analysis, conserved phosphosite analysis, and Motif-X searches

A B-fabric (PostgreSQL, <http://www.postgresql.org/>) database was queried and all identifications were loaded into Microsoft Excel 2011 (Microsoft Corp., <http://www.microsoft.com>) and analyzed with different pivot tables to answer specific questions. We compared our phosphopeptide and phosphoprotein identifications with the PhosPhAt, a database of all empirically identified phosphoproteins of *Arabidopsis* (as of 17/05/2011; data kindly provided by Waltraud Schulze) (Holmes-Davis *et al.*, 2005; Durek *et al.*, 2010; Heazlewood *et al.*, 2008) and identified a subset of peptides and proteins that were still missing from the database as none of the large-scale studies compiled in the database used pollen material.

For these pollen-enriched phosphopeptides and phosphoproteins we did a more detailed analyses using the TopGO algorithm to check for over-represented gene ontology categories (Alexa *et al.*, 2006). As a list of foreground proteins we used the list of pollen-specific phosphoproteins (not present in PhosPhAt) and compared it to 3 distinct backgrounds: the pollen proteome (protein list extracted from Pep2Pro containing the latest TAIR10 annotations) (Grobei *et al.*, 2009; Baerenfaller *et al.*, 2011), the total empirical proteome (Pep2Pro, an *Arabidopsis* tissue-specific proteome database) (Baerenfaller *et al.*, 2011), and the total empirical *Arabidopsis* phosphoproteome (PhosPhAt) (Holmes-Davis *et al.*, 2005; Durek *et al.*, 2010; Heazlewood *et al.*, 2008).

To determine whether phosphosites in *Arabidopsis* proteins essential to pollen development may have a functional role, we investigated whether the sequence contexts of the phosphosites were conserved in the orthologous proteins of *Vitis vinifera*, *Populus trichocarpa*, *Zea mays*, and *Oryza sativa* was conserved. We identified the orthologous proteins using NCBI’s UniGene Database (<http://www.ncbi.nlm.nih.gov/UniGene/>), and then aligned orthologous protein sequences to

determine whether the sequence context of a phosphosite was conserved using NCBI's COBALT constraint-based multiple protein alignment tool (<http://www.ncbi.nlm.nih.gov/tools/cobalt/>).

To identify over-represented sequence motifs among the phosphosite-peptides we used pollen-specific phosphopeptides with confirmed phosphosites. The analysis was conducted using only Ser phosphorylated peptides as Thr and Tyr phosphorylated peptides were less abundant. A total of 220 Ser phosphorylated peptides were used for extracting the over-represented motifs with a significance value of  $P=0.000001$ .

#### 4.1.1.6 Genome wide expression profiles of pollen-specific phosphoproteins

Original ATH1-array data from different *Arabidopsis* tissues were downloaded from public repositories to generate an atlas of expression of different tissues and cell types (see Supplemental Table 1 for details), and included data from discrete seed compartments (Le *et al.*, 2010; Array Express dataset GSE12404, [www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)), cell types of the shoot apical meristem (Yadav *et al.*, 2009), and a dataset profiling early stages of microgametogenesis (Honys and Twell, 2004), sperm and mature pollen (Borges *et al.*, 2008), female gametophytic cell types (Wuest *et al.*, 2010), root cell types (Brady *et al.*, 2007), and various other tissues (Schmid *et al.*, 2005). This tissue atlas totally includes a set of 63 tissue types of gametophytic, sporophytic, and embryonic origin. Gene expression signals were calculated using dCHIP software (<http://biosun1.harvard.edu/complab/dchip/>, version 2010) with invariant-set normalization and a PM-only model. Probeset definitions were used as described (Schmidt *et al.*, 2011). dCHIP expression estimates were imported into the R statistical software, which was used for all further data processing. Heatmaps were generated using functionality provided by the R-package gplots ([www.bioconductor.org](http://www.bioconductor.org), Version 2.8.0).

## 4.2 References

- Alexa A, Rahnenfuhrer J, Lengauer T** (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600-1607
- Baerenfaller K, Hirsch-Hoffmann M, Svozil J, Hull R, Russenberger D, Bischof S, Lu Q, Gruissem W, Baginsky S** pep2pro: a new tool for comprehensive proteome data analysis to reveal information about organ-specific proteomes in *Arabidopsis thaliana*. *Integr Biol (Camb)* **3**: 225-237
- Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP** (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**: 1285-1292
- Berndt P, Hobohm U, Langen H** (1999) Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis* **20**: 3521-3526
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN** (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801-806
- Durek P, Schmidt R, Heazlewood JL, Jones A, MacLean D, Nagel A, Kersten B, Schulze WX** (2010) PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res* **38**: D828-834
- Elias JE, Gibbons FD, King OD, Roth FP, Gygi SP** (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat Biotechnol* **22**: 214-219
- Grobei MA, Qeli E, Brunner E, Rehrauer H, Zhang R, Roschitzki B, Basler K, Ahrens CH, Grossniklaus U** (2009) Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Res* **19**: 1786-1800
- Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, Schulze WX** (2008) PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res* **36**: D1015-1021
- Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S** (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics* **5**: 4864-4884
- Hony D, Twell D** (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol* **5**: R85



- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS** (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551-3567
- Qeli E, Ahrens CH** (2010) PeptideClassifier for protein inference and targeted quantitative proteomics. *Nat Biotechnol* **28**: 647-650
- Reiland S, Messerli G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, Gruissem W, Baginsky S** (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol* **150**: 889-903
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU** (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501-506
- Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U** (2011) Transcriptome analysis of the *Arabidopsis* megaspore mother cell uncovers the importance of RNA helicases for plant germline development. *PLoS Biol* **9**: e1001155
- Turker C, Akal F, Schlapbach R** (2011) Life sciences data and application integration with B-fabric. *J Integr Bioinform* **8**: 159
- Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer J, von Mering C, Grossniklaus U** (2010) *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol* **20**: 506-512

## 4.3 Supporting Figure and Table Legends

### 4.3.1 Supporting Table S1: Overview of Pollen Phosphoproteome

The Microsoft Excel list shows all the phosphoproteins and phosphopeptides identified from *Arabidopsis* pollen grains. Relevant information on the sub-fraction of protein digests and chromatography techniques used to enrich a particular phosphopeptide is also listed. The sheet also describes the confirmed phosphorylation sites (called "Phosphosite-peptides") on 609 phosphopeptides. PeptidesClassifier mapping of each peptide into different classes (1A, 1B, 2A, 2B, 3A, 3B) are also listed. Finally we provide the transcriptomic evidence for each phosphopeptide, confirming that the phosphopeptides were correctly identified.

Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

### 4.3.2 Supporting Table S2: List of Phosphoproteins Required for Normal Pollen Development

This Microsoft Excel list contains *Arabidopsis* genes with observed phenotypes at late or early stages of pollen development resulting from mutations in the corresponding gene. The table contains the gene name, the corresponding *Arabidopsis* gene identifier (AGI), the respective literature reference, and the identified phosphosites (indicated by \*) or phosphopeptides (bold). The corresponding protein identifiers from NCBI are indicated for *A. thaliana* and, if available, for *V. vinifera*, *P. trichocarpa*, *Z. mays*, and *O. sativa*; conservation of the amino acid residue in the sequence context of the phosphosite is indicated (amino acid position within the predicted protein, A.A.).

Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

### 4.3.3 Supporting Table S3: List of Published Arabidopsis Pollen Mutants (15.09.2011).

The row with colored lines displays the pollen mutants for which a phosphoprotein was identified in our analysis.

Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

### 4.3.4 Supporting Table S4: Pollen-enriched phosphoproteins.

The table lists the phosphoproteins that were exclusively identified in our study and are missing from the large-scale phosphoproteome database (PhosPhAt, dated 17.05.2011; kindly provided by Waltraud Schulze).



Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

#### 4.3.5 Supporting Table S5: List of Kinases Found in the Pollen Phosphoproteome.

Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

#### 4.3.6 Supporting Table S6: List of RLKs Found in the Pollen Phosphoproteome.

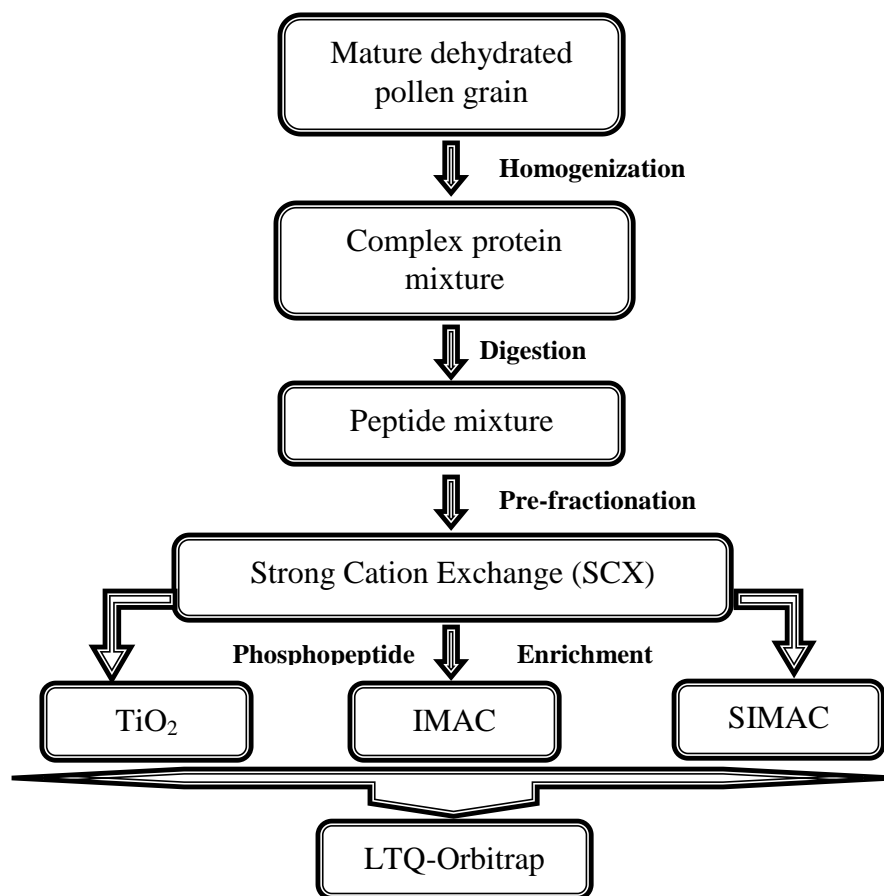
Interestingly, out of 46 subfamilies of RLKs only four were represented in our pollen phosphoproteome. These four families are represented by nine phosphopeptides. Seven members of RLK subfamilies were identified by information-rich peptides (class 1A and 2B). However, the last subfamily was represented only by class 3B peptides. Respective transcriptomic evidence for each subfamily is provided (based on Honys and Twell, 2004).

Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2012.05061.x/supinfo>

#### 4.3.7 Supporting Figure S1: Outline of Experimental Procedures.

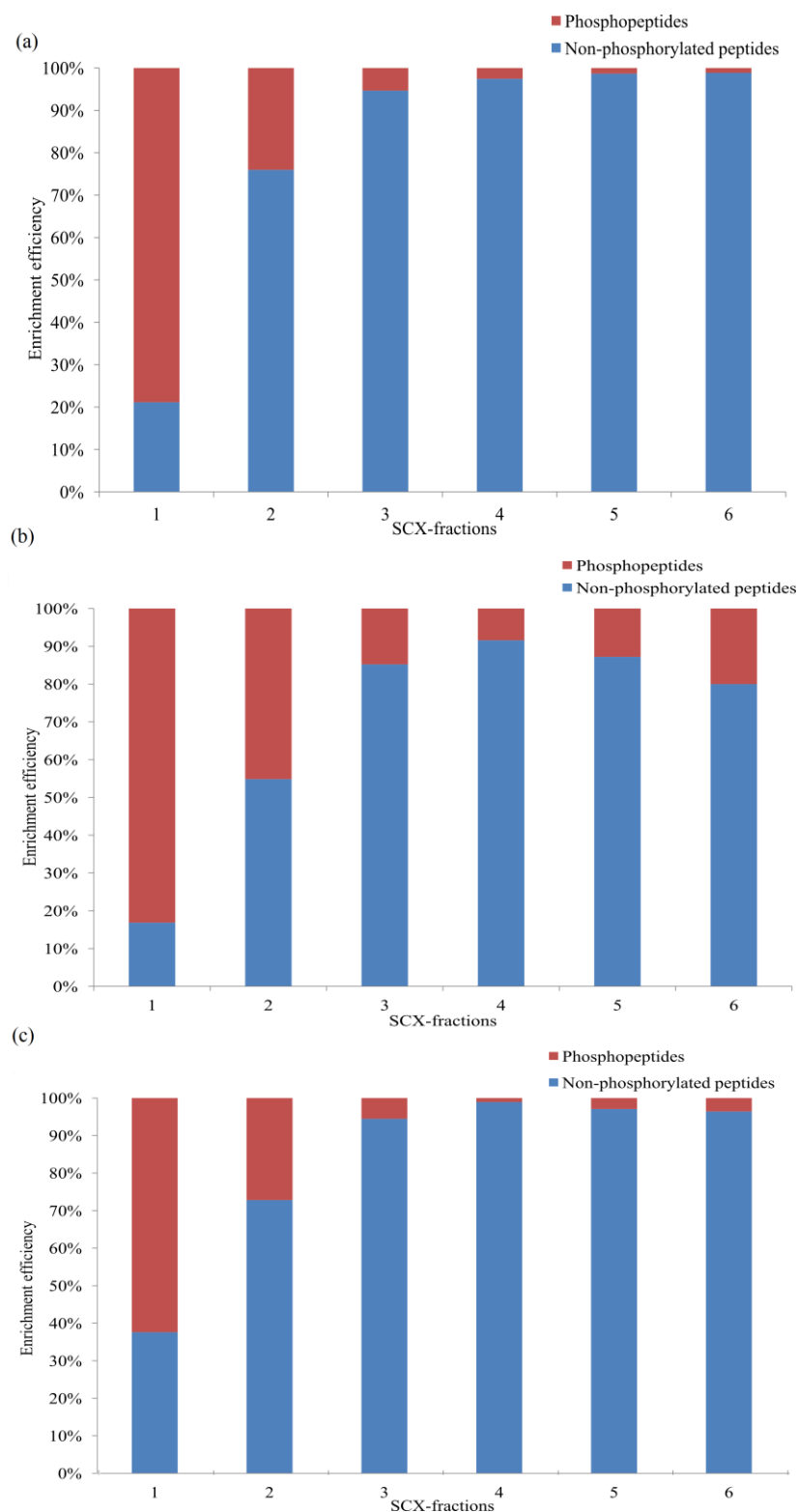
Work flow for the acquisition of the phosphopeptides from *Arabidopsis* mature dehydrated pollen grain. Comprehensive coverage was obtained using three different enrichment chromatographies (TiO<sub>2</sub>: Titanium dioxide, IMAC: Immobilized Metal ion Affinity Chromatography, SIMAC: Sequential elution from IMAC).

(a) Immobilized Metal ion Affinity Chromatography (IMAC), (b) Titanium dioxide (TiO<sub>2</sub>) chromatography, and (c) Sequential elution after IMAC (SIMAC).



**Supplemental figure S1.** Outline of experimental procedures.

#### 4.3.8 Supporting Figure S2: Representation of phosphopeptide enrichment in six different SCX fractions coupled with three different phospho-enrichment chromatographies.



**Supplemental figure S2:** Representation of phosphopeptide enrichment in six different SCX fractions coupled with three different phospho-enrichment chromatographies

## Phosphoproteome map of pollination specific tissue involved in pollen – stigma interaction in *Arabidopsis thaliana*

---

**Pururawa Mayank**<sup>1, 2, \*</sup>, Jonas Grossman<sup>3</sup>, Samuel E.Wuest<sup>4</sup>, Paolo Nanni<sup>3</sup>, Ueli Grossniklaus<sup>1, 2, \*</sup>

<sup>1</sup>Centre for Model Organism Proteomes, University of Zürich\*, Zürich, Switzerland

<sup>2</sup>Institute of Plant Biology & Zürich-Basel Plant Science Centre, University of Zürich\*, Zürich, Switzerland

<sup>3</sup>Functional Genomics Center Zürich, University of Zürich\* and ETHZ\*, Zürich, Switzerland

<sup>4</sup>Institute of Evolutionary Biology and Environmental Studies & Zürich-Basel Plant Science Centre, University of Zürich\*, Zürich, Switzerland.

\*For correspondence (fax +41 446348204; e-mail [grossnik@botinst.uzh.ch](mailto:grossnik@botinst.uzh.ch)).

\*Street address: Zollikerstrasse,107, Zurich, Switzerland

## 5.1 Summary

In angiosperms, sexual reproduction initiates with pollination, when the pollen grain (male gametophyte) is transferred to receptive tissue of the pistil, the stigma. Upon arrival on the stigmatic surface the pollen adheres, hydrates, germinates and grows a pollen tube that later delivers the two sperm cells to the embryo sac (female gametophyte; FG) to effect double fertilization. Various studies have shown that mature dehydrated pollen grains contain most of the required transcripts and proteins for the initial step of pollen tube (PT) growth and development. It is therefore important to explore the role of post-translational modifications (PTMs) in the pollination process where cell-cell communication event takes place and rapid cellular responses occur. Here, we focus on protein phosphorylation and report a phosphoproteomics study conducted on *Arabidopsis* pollination-specific tissues by comparing the phosphoproteomes of the unpollinated and pollinated male and female tissues involved. We identified 1985 unique phosphopeptides isolated from pollen, unpollinated gynocia, and 1hap gynocia (one hour after pollination stigma and ovary individually). These phosphopeptides correspond to 1198 unique phosphoproteins with a false discovery rate of 0.64%. 946 phosphorylation sites were mapped using a heuristic approach. Among the 946 unique phosphopeptides, we found only one Tyr-phosphorylated peptide [VTSEDFMT\*EY\*VVTR; AT2G43790]. In comparison to our previous phosphoproteomics study of mature *Arabidopsis* pollen we identified an additional 607 phosphopeptides corresponding to 414 phosphoproteins uniquely identified in the current study. Using a label-free approach (PROGENESIS LC-MS) we quantified 1450 phosphopeptides that change in their abundance between the unpollinated and pollinated tissues that could be key regulators of pollen-stigma interactions during self-compatible species.

## 5.2 Introduction

The angiosperms life cycle alternates between a diploid sporophytic and a haploid gametophytic generation. The gametophytes, which produce the male and female gametes (sperm cells, egg and central cell) respectively play a central role in fertilization. The mature male gametophyte (pollen grain) consists of a large vegetative cell and two sperm cells that are enclosed in its cytoplasm (Holmes-Davis *et al.*, 2005). The female gametophyte (embryo sac) consists of seven cells: three antipodal cells, two synergid cells, and two female gametes, the central cell and the egg cell. The embryo sac is embedded within the ovule, which in turn is enclosed by sporophytic tissues that form the gynoecium (carpel or pistil) (Shi and Yang, 2011). In terrestrial plants, the communication between the haploid pollen and diploid gynoecium is initiated at the stigma, a receptive surface at the tip of gynoecium. The pollen grain adheres to the stigma and, upon recognition, hydrates and germinates producing a pollen tube, which elongates and penetrates the stigma, grows through style and transmitting tract, to carry the sperm cells to the female gametophyte, ensuring successful double fertilization (Dresselhaus, 2006). These pollen-pistil interactions are characterized by a multitude of cell-cell communications and signaling processes (Hiscock and Allen, 2008). The process culminates in double fertilization, which results in the formation of a diploid zygote and a triploid endosperm initiating seed development.

In flowering plants, whether self-compatible or self-incompatible, the pollen grains are dispersed from the anther and transferred to the stigma by pollinators, wind, or directly in case of self-fertile plants like *Arabidopsis*. The *Arabidopsis* mature pollen grain represents a highly simplified organism that contains a complex array of transcripts and proteins required for germination and rapid pollen tube growth (Honys and Twell, 2004; Holmes-Davis *et al.*, 2005; Pina *et al.*, 2005; Wang *et al.*, 2008; Grobei *et al.*, 2009; Borges *et al.*, 2008; Mayank *et al.*, 2012). As germination and initial pollen tube growth doesn't require protein synthesis, some of the stored proteins must undergo activation upon pollination, most likely by post-translational modification (PTM). In particular, phosphorylation or dephosphorylation may be involved in the initial cell-cell communication and signaling processes during pollination. Phosphoproteins analysis of mature dehydrated pollen grains showed a significant enrichment in processes such as "pollen germination," "pollen tube development," "fertilization," and "male gametes generation" suggesting a key role of phosphorylation in germination and initial pollen tube growth (Mayank *et al.*, 2012). This study provides a framework for understanding the dynamics of protein phosphorylation in male and female tissue upon pollination.

In most angiosperms the pollen is dispersed from the anther in a dehydrated state. Therefore, immediately after landing on the stigmatic surface, water is supplemented by the stigmatic cells (Gaude, and McCormick, 1999), allowing the pollen grain to hydrate and initiate cell-cell interactions that trigger germination (Dickinson, 1986). After germination pollen tube penetrates the stigmatic cell wall and grows rapidly towards the style. Stigmas of angiosperms can be broadly classified into two categories, *i.e.* wet or dry, based on their surface secretion. In wet-stigma species, such as tobacco, petunia and lilly pollen capture and adhesion by the stigmatic secretion appears to be non-specific and pollen hydration is largely unregulated (Heslop-Harrison & Shivana, 1977; Hiscock and Allen, 2008). In contrast, the studies conducted on dry-stigma species, mostly from the *Brassicaceae* such as *Arabidopsis* and rapeseed suggest a high species-specificity for pollen adhesion, and hydration which are tightly regulated (Dickinson, 1995; Hiscock and Allen, 2008).

Species of the *Brassicaceae* have been used as a model system to understand the physiological and molecular mechanisms underlying self-incompatible pollen-stigma interactions, where self-recognition blocks pollen hydration (Dickinson, 1995; Dickinson, *et al.*, 1998; Hiscock and Allen, 2008; Chapman and Goring, 2010). However, knowledge about the signaling events during pollination in self-compatible species such as *Arabidopsis thaliana* is still limited. Previous transcriptome and proteome studies of the pollen (Grobei *et al.*, 2009; Honys and Twell, 2004; Holmes-Davis *et al.*, 2005; Mayank *et al.*, 2012), stigma (Swanson *et al.*, 2005), and pistil (Scutt *et al.*, 2003) have failed to provide a framework for the molecular events taking place during pollen-stigma interactions. Therefore this issue was addressed in a study of the kinetics of gene expression over several time points, during the pollination process (Boavida *et al.*, 2011). This provided a view of transcriptional gene regulation underlying sexual plant reproduction. However, the post-transcriptional changes involved in pollination are still unknown. The stigmatic responses regulating pollen adhesion, recognition, hydration, and germination are likely depending on conformational changes of some proteins from either sexual partner. These changes are mediated by PTMs such as phosphorylation and dephosphorylation. Therefore, we isolated, identified, and quantified phosphoproteins and phosphopeptides from pollination-specific tissues pre- and post-pollination.

As discussed in a previous study the phosphoproteome of pollen grain contains a large set of phosphorylated proteins that might be subjected to changes in their phosphorylation status or abundance upon pollination (Mayank *et al.*, 2012). However, the mere identification of phosphoproteins in male and female pollination-specific tissues will not provide insights into the protein phosphorylation dynamics involved in the signaling cascades during pollination. To

address this question we conducted label-free quantification of pollination-specific phosphoproteins and identified candidates which might be involved in the pollen-stigma interaction process.

We identified a total of 1985 unique phosphopeptides, which were isolated from pollen, unpollinated gynoecia (UP) and gynoecia at one hour after pollination (1HAP) gynoecium (stigma and ovary separately). This corresponds to 1198 unique phosphoproteins, with a false discovery rate of 0.64% at the spectral level. Additionally, we quantitated 1450 phosphopeptides that change in their abundance or phosphorylation status during pollination. A comparative study between our pollination-specific phosphoproteome and previous transcriptome and proteome studies (Boavida *et al.*, 2011), (Qin *et al.*, 2009), (Baerenfaller *et al.*, 2008) revealed that 93% of 1HAP and 88.3% of UP gynoecium phosphoproteins have corresponding transcripts or proteins, validating the quality of our dataset. Candidate showing differential abundance between unpollinated and pollinated tissue can increase our understanding on the signaling event that regulate the pollen tube biology *in vivo*.

In conclusion, this study provides a framework for post-transcriptional regulation involved in pollen-stigma interactions in a self-compatible species and will enable a better understanding of the signaling events that take place upon pollination.

## 5.3 Results and Discussion

### 5.3.1 An extended atlas of the pollen phosphoproteome identifies 414 novel phosphoproteins

In order to study the differential abundance of phosphopeptides/phosphoproteins in pollination-specific tissues, it was important to analyze all pollination-specific tissues (pollen, stigma and ovary) simultaneously to avoid technical variability amongst the samples. Regardless of which label-free approach is used, it is very important that the sample preparation and sample injection steps are done in parallel to minimize the variation in the experiment. The Progenesis LC-MS label-free quantitative tool was used to establish the relative quantification of phosphopeptides present in the different reproductive tissues analyzed.

Although we previously characterized the phosphoproteome of mature *Arabidopsis* pollen grains (Mayank *et al.*, 2012), it was important to re-analyze pollen grains along with other pollination-specific tissues in order to generate a label-free quantification map of the phosphopeptides from post-pollination tissue. Phosphoproteins from pollen and unpollinated

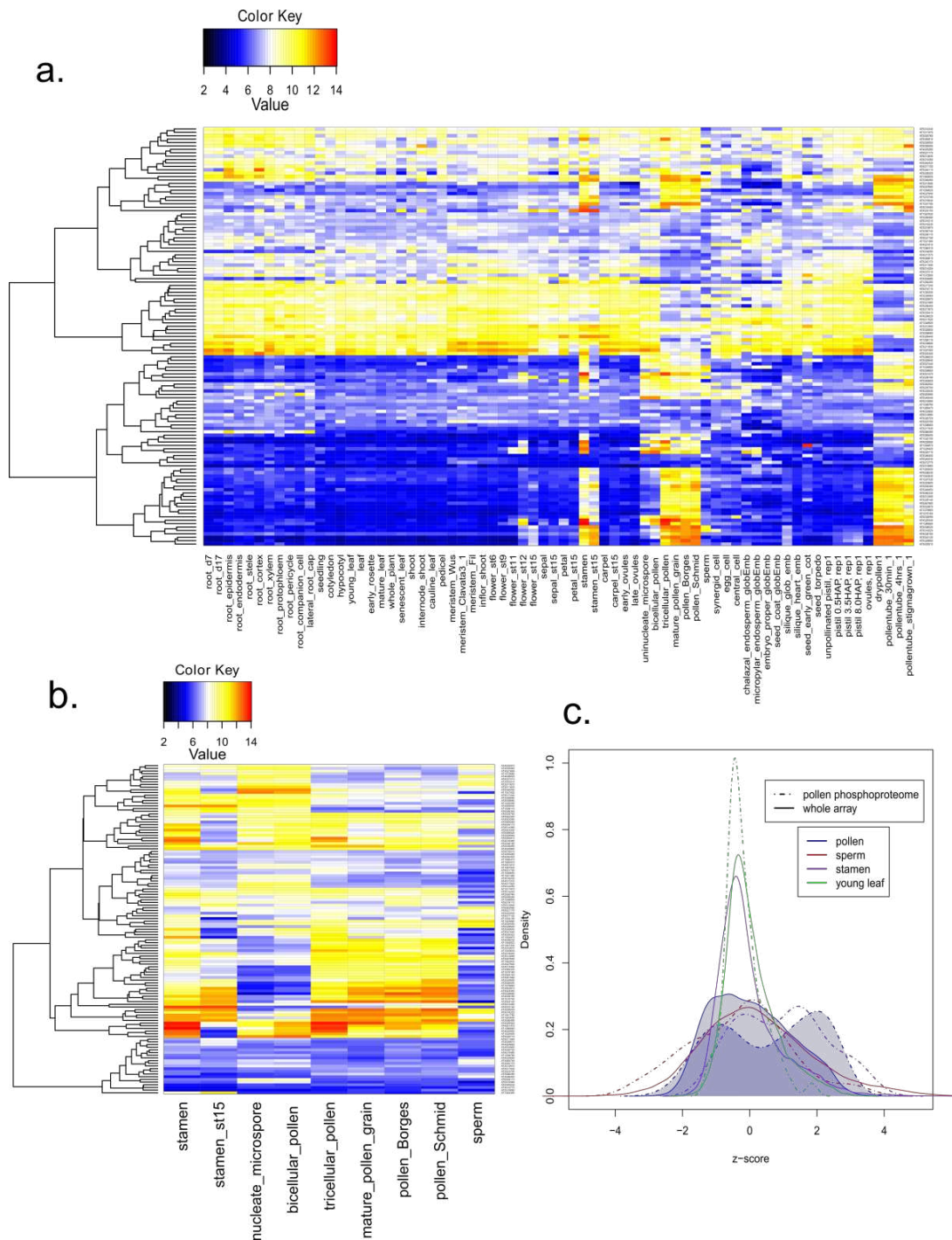


female tissues were used as a reference atlas to generate and quantify the phosphoproteome of tissue 1hap. using our previously described protocol in (Mayank *et al.*, 2012), a total of 995 phosphopeptides were identified that correspond to 831 phosphoproteins. 607 phosphopeptides corresponding to 414 phosphoproteins we uniquely identified in the current study (Table 4.1).

Tissue	Specific Phosphopeptides	Specific Phosphoproteins
Pollen Old (a)	607	200
Pollen New	607	414
Shared	388	417

**Table 5.1:** Comparison between our previous study on *Arabidopsis* mature pollen phosphoproteome and the current study. In previous study (Mayank *et al.*, 2012). (a) (Mayank *et al.*, 2012)

A comparative study with the PhosPhAT database (phosphorylation database of *Arabidopsis*) identified a new subsets of that was absent from the database. We checked the expression of the genes encoding these newly identified phosphoproteins in an expression atlas containing a large number of *Arabidopsis* tissues and cell types (Wuest *et al.*, 2010), including the transcriptomes of various stages of male gametophyte development (Honys and Twell, 2004), the pollen tube transcriptome (Y., Wang *et al.*, 2008), and the transcriptome of a time course after pollination (Boavida *et al.*, 2011). We found that the majority of these genes show a relatively high expression in pollen at various stages or in post-pollinated gynoecia (0.5, 3.5, and 8 HAP). However, in contrast to the previous study, we found that a substantial fraction of these genes exhibited a low expression in mature pollen while high expression is seen only during early stages of pollen development, and for sporophytic tissues (Figure 5.1 a-c). Since phosphopeptides from sporophytic tissues have been well characterized in previous studies (Arsova and Schulze, 2012), we expected nearly-exclusive gene expression at mature pollen stages. A possible explanation for this discrepancy could be that we identified proteins exhibiting pollen-specific phosphorylation, a hypothesis that needs to be tested in future experiments.



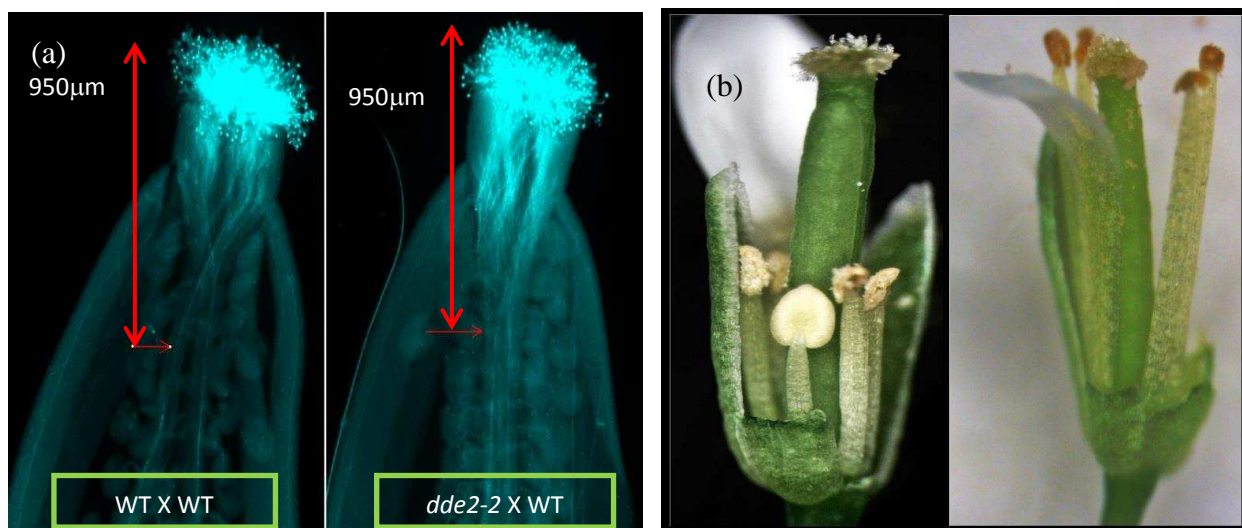
**Figure 5.1:** Newly identified pollen phosphoproteins expression using AtPANP analysis: **(a)** Phosphoproteins that are absent from PhosPhAt, which include our previous study (Mayank *et al.*, 2012). Representation of relative expression values (per gene- normalised; also termed z-scores) across the tissue atlas (Wuest *et al.*, 2010) for genes encoding phosphoproteins that are absent from PhosPhAt; black and blue show low expression and yellow/red show highest expression. **(b)** Representation of relative expression values (row-scaled; also termed z-scores) across the tissue atlas, with a focus on male reproductive tissues/cells. **(c)** Overall z-scores are shifted towards high values indicating predominant expression of genes in male reproductive tissues/cells, particularly in mature pollen.

### 5.3.2 Qualitative phosphoproteome of pollination-specific tissue

Both in self-compatible or self-incompatible species, pollination require interaction between the pollen (haploid male gametophyte) and the stigma (a tissue of diploid female sporophyte). Pollen-stigma interactions are highly specific and tightly regulated by the stigma. Moreover, the

pollen coat (exine) also contains molecules involved in the interaction with the stigma (Lord, 2003). However, in self-incompatible species inbreeding is prevented by the S-haplotype. The male determinant of SI in *Brassicaceae*, is a small cysteine-rich protein (SCR), which is recognized by the female S-locus-related glycoprotein (SLR) (Doughty, 2000; Kachroo *et al.*, 2001; Schopfer, 1999; Shiba, 2001) and arrests the pollen germination either before or after pollination.

In order to identify and quantify pollination-specific phosphoproteins and phosphopeptides, we use the *delayed-dehiscence2-2* (*dde2-2*) mutant as the female parent. *dde2-2* is defective in anther dehiscence and filament elongation (Von Malek *et al.*, 2002) (Figure 5.2 b), and, thus, exhibits a male-sterile phenotype. This greatly facilitates manual pollination that is required for the collection of a large number of pollinated pistils, a critical step in a phosphoproteomics study. In order to test whether pollination and pollen tube growth is similar to the wild type in *dde2-2* mutants, we characterized pollen tube growth and seed set in *dde2-2* crossed to wild-type pollen (Figure 5.2 a). Using aniline blue staining to trace the pollen tube growth, no obvious differences were observed between *dde2-2* and wild-type controls. Seed set in the two genetic backgrounds did also not differ significantly (Supplementary table S4). Thus, *dde2-2* mutant displayed complete female fertility and normal pollen tube growth when crossed with wild type pollen. In order to characterize the phosphoproteome of pollinated pistils, we collected a large number of gynoecia unpollinated and pollinated (approximately 1000). Protein extraction was performed using a previously established protocol (Mayank *et al.*, 2012).



**Figure 5.2:** Comparison of pollination in an pollen tube growth Col WT and the *dde2-2* mutant **(a)** Aniline blue staining of 5h ap (hour after pollination) showing normal PT growth when back crossed with wild-type pollen in *dde2-2* mutant and wild type plants. **(b)** *A. thaliana dde2-2* mutant (left) is a male sterile mutant due to a mutation in AOS gene, which encodes a key enzyme of jasmonic acid (JA) biosynthesis. Therefore, the fertility in mutant plant can be restored by application of methyl jasmonate. *dde2-2* is defective in anther dehiscence and filament elongation because of which it is self-infertile and facilitates manual pollination.

The extracted proteins from pollen and unpollinated and pollinated female *dde2-2* (stigma and ovary) were pre-fractionated using strong cation exchange (SCX) and enriched using immobilized metal ion affinity chromatography (IMAC) before being loaded onto a LTQ Orbitrap coupled with the nano high pressure liquid chromatography (HPLC) system. The resulting raw files were then converted to mascot generic files (mgf) and analyzed by MASCOT ([www.matrixscience.com](http://www.matrixscience.com)). The results were uploaded to MS-parser and were analyzed using Microsoft Excel and Pivot tables. We identified a total of 1985 unique phosphopeptides isolated from pollen, unpollinated (UP) and pollinated gynoecia (1HAP ovary and stigma individually) which corresponded to 1198 unique phosphoproteins with a false discovery rate (FDR) of 0.64% (at the spectra level) (Supplement table S1). All the phosphopeptides reported in this study are above a Mascot ion score  $\geq 20$  and are significant top ranking (bold red) peptides. To increase the quality of the data and decrease the FDR we manually validated the entire single hit spectra (129) as previously described (Mayank *et al.*, 2012). To extract the maximum information from our dataset, we used a deterministic peptide classification scheme, Peptide Classifier (Qeli and Ahrens, 2010), which classified 93% of the phosphopeptides into high information content classes (1a, 1b, 2a and 2b) for which corresponding protein sequences and gene loci were unambiguously identified. 946 phosphorylation sites were identified using a heuristic approach as previously described (Supplementary table S4) (Mayank *et al.*, 2012). Among the 946 unique phosphopeptides, we could only validate a single Tyr-phosphorylated peptide [VTSESDFMT\*EY\*VVTR; AT2G43790], which belongs to Peptide Classifier class 1a. A total of 2214 singly phosphorylated, 189 doubly phosphorylated and 4 multiply phosphorylated peptides were found. Occurrences of these sites of phosphorylation was 90.8% on Ser, 8.8% Thr and 0.4% on Tyr, which is comparable to the previous phosphoproteomics studies (Reiland *et al.*, 2009; Mayank *et al.*, 2012).

### 5.3.3 Over-represented GO categories from *dde2-2* gynoecia 1HAP, have obvious role in pollen-stigma interaction

The gynoecium of *Arabidopsis* has dry stigma, which shows a high specificity for pollen adhesion and regulates its hydration both temporally and spatially by supplying water before the germination of pollen tube. The papillar cells forming the stigmatic surface secrete esterases that act on the pollen coat and reorganize it, both structurally and physiologically, so that the stigma can supply water to the pollen for its hydration (Doughty *et al.*, 1993). Pollen adhesion on the stigmatic surface lasts for 10-15 mins upon its arrival and can greatly vary, depending on the maturity and thickness of SPC (Dickinson, 1995). Successful adhesion and hydration is

facilitated by the translocation of proteins along with water between the partners, which in turn may itself be stimulated by pollination. It has been observed that when stigmatic cells are ablated with the use of the diphtheria toxin A chain (DT-A), a potent inhibitor of translation, the pollen viability and development is affected and pollen tube growth is arrested in both *Brassica* and *Arabidopsis* (Sarker *et al.*, 1988; Kandasamy *et al.*, 1993). However, impairment of S-locus glycoproteins (SGP; high water-trapping capacity) which are conserved in *Brassica* species, do not affect the receptiveness of the papillar cells and pollen tube growth although papillar cell were biochemically impaired by using of phosphatase inhibitors (Kandasamy *et al.*, 1993). Apparently, dephosphorylation plays an important role in *Brassica* self-incompatible species, where protein phosphatase inhibitors disrupt the pollen recognition system arresting pollen hydration (Sarker *et al.*, 1988; Kandasamy *et al.*, 1993; Dickinson, 1995). However, it is not known whether phosphorylation events also regulate pollen-stigmas interaction in *Arabidopsis*. Hence, we carried out a phosphoproteomics study of gynoecia which we separated into stigma and ovary to enable a better understanding of the biology underlying successful pollination in *Arabidopsis*.

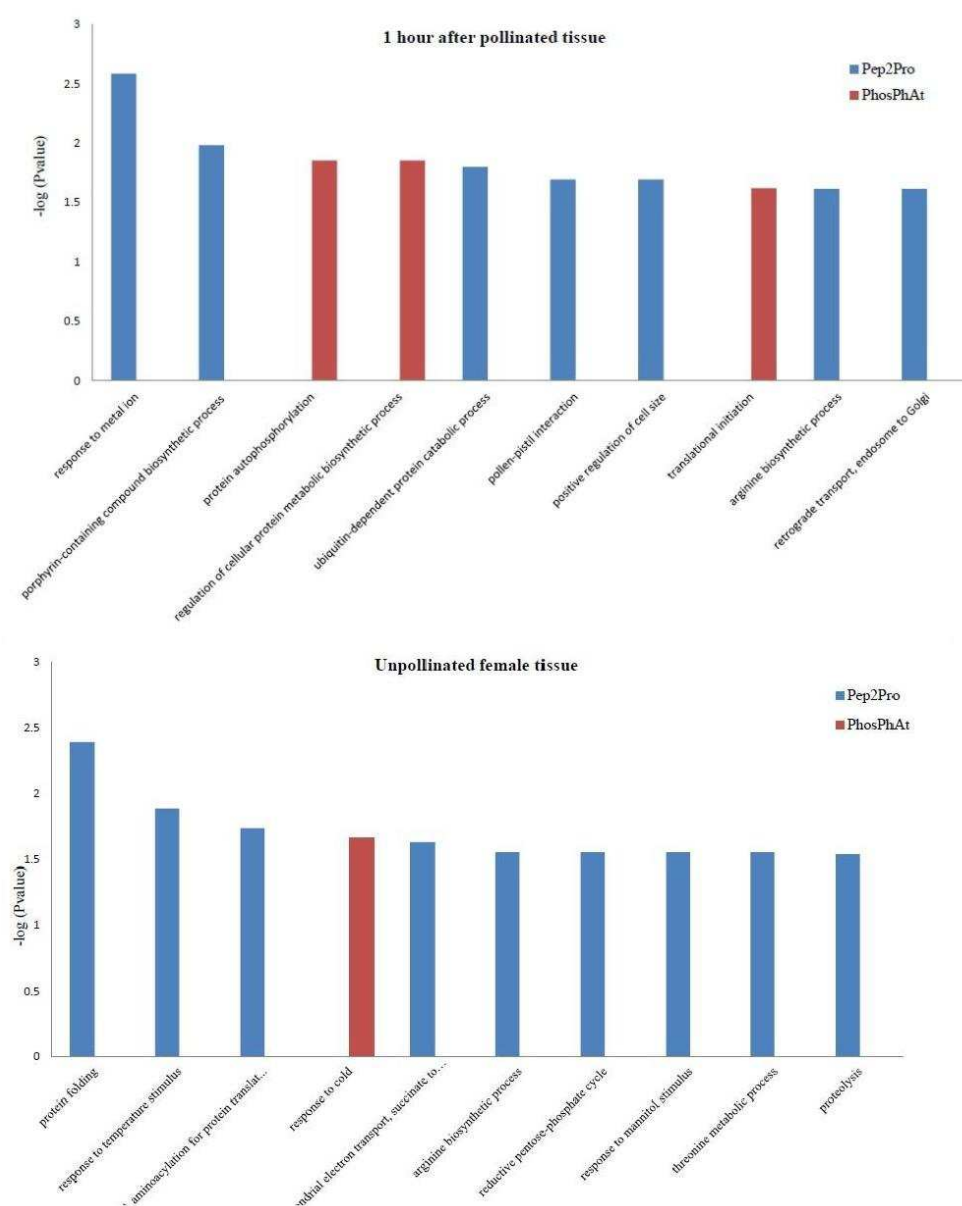
As previously stated, we performed the phosphoproteome analyses of gynocia which exhibits a male sterile phenotype Figure 5.2 (b). The need for a large number of unpollinated gynocia and gynocia 1HAP was met by using *dde2-2* which allows manual pollination. We successfully isolated 444 unique phosphopeptides, corresponding to 261 unique phosphoproteins from unpollinated gynocia and gynocia 1HAP pollinated female tissue. The number of phosphopeptides and the corresponding phosphoproteins from individual tissues are listed in (Table 4.2). Most of the identified phosphoproteins are absent in the PhosPhAt database, presumably because no similar phosphoproteomics study has been reported before.

Tissue	Unique & Specific phosphopeptides	Unique & Specific phosphoproteins
Unpollinated stigma	71	46
Unpollinated Pistil	225	121
1HAP stigma	48	37
1HAP pistil	100	56
Pollen	995	478
Common to all tissues	546	460

**Table 4.2:** List of the unique set of identified phosphopeptides and corresponding phosphoproteins from pollination-specific tissues.

To characterize the phosphoproteins based on their molecular function (MF), we performed gene ontology (GO) studies separately on phosphoproteins extracted from unpollinated gynocia and on gynocia 1HAP, using the TopGo algorithm (Alexa *et al.*, 2006; Mayank *et al.*,

2012). The analysis was performed by using two different foregrounds (84 phosphoproteins from unpollinated gynoecia and 67 phosphoproteins from gynoecia 1HAP; that were absent in the PhosPhAt database) against two distinct backgrounds; the total empirical proteome (Pep2Pro, an *Arabidopsis* tissue-specific proteome database (Hirsch-Hoffmann *et al.*, 2012; Baerenfaller *et al.*, 2011) and PhosPhAt (Heazlewood *et al.*, 2008; Durek *et al.*, 2010; Arsova and Schulze, 2012). GO categories were considered over-represented based on a p-value (0.05) (Figure 5.3). We compared the top ten over-represented categories from both foregrounds. Interestingly, gynoecium 1HAP tissue showed an over-representation of GO categories that are related to pollination, such as “protein autophosphorylation,” “pollen-pistil interaction,” and “positive regulation of cell size,” “arginine biosynthetic process,” and “translational initiation. Many of these categories have an obvious role in pollen-stigma interaction.



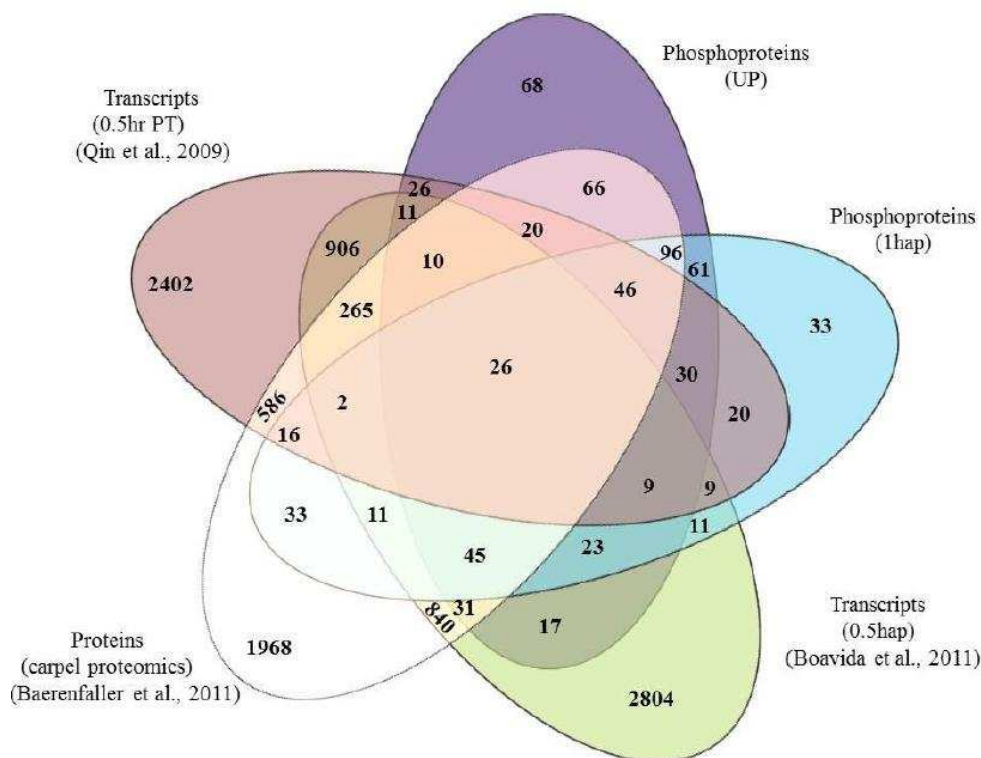
**Figure 5.3:** TopGO analysis on unpollinated gynoecia and gynoecia 1HAP: The analysis was performed using two backgrounds Pep2Pro (Baerenfaller *et al.*, 2011) and PhosPhAt (Heazlewood *et al.*, 2008) (a) Over-represented



categories in gynoecia 1hap have obvious role in pollination (b) unpollinated gynoecia over-represented GO categories.

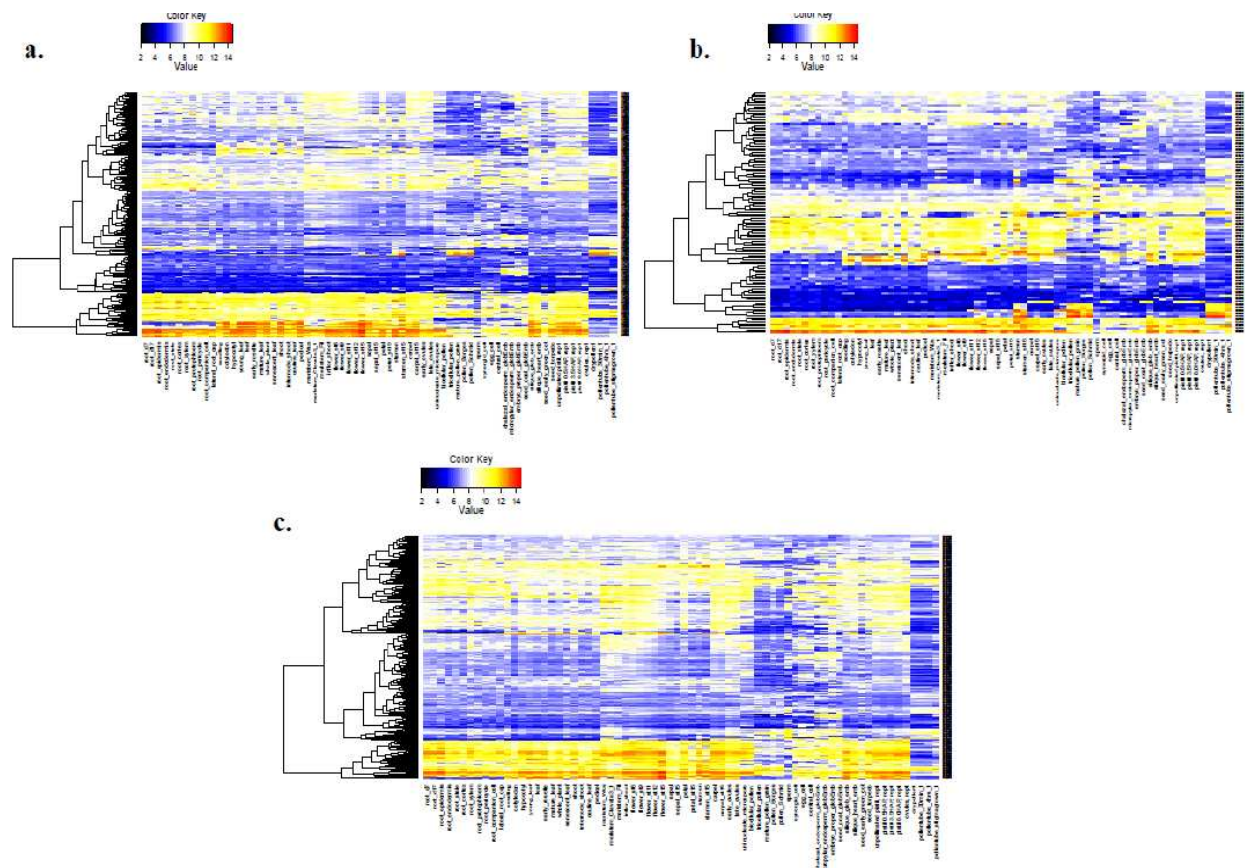
As pollen contains most of the required transcripts and proteins for hydration and germination, the interaction with the papillar cells causes some of these proteins to translocate or undergo a conformational change mediated by PTMs, for instance phosphorylation. The GO term “Pollen-stil interaction” was over-represented in our GO analysis suggesting that this interaction might be driven by a series of phosphorylation events. “Translational initiation” facilitates the synthesis of proteins from stored mRNA upon hydration. Finally the term “positive regulation of cell size” is over-represented in our dataset, related to the rapid increase in size of the pollen during pollen tube growth following germination.

To validate the identified phosphoproteins we compared the datasets from pollinated and unpollinated *dde2-2* gynoecia with published data for the transcriptome of 0.5 HAP pistils; (Boavida *et al.*, 2011), of pollen tubes grown *in vitro* for 0.5hrs (Qin *et al.*, 2009) and proteome of carpels (Baerenfaller *et al.*, 2008) (Figure 5.4; Supplementary Table S2). Interestingly, this comparison revealed that 88.3% and 93% of phosphoproteins from unpollinated gynoecia and gynoecia 1HAP respectively, have corresponding transcripts or proteins which have been detected in these previous studies, thus validating the quality of our dataset.



**Figure 5.4:** Overlap of unpollinated gynoecia and gynoecia 1HAP with previous transcriptomics and proteomics datasets; the Venn diagram visualizes the overlap of unpollinated gynoecia and gynoecia 1HAP with transcriptomics (0.5 hap; Boavida *et al.*, 2011; 0.5h *in vitro* pollen tube; Qin *et al.*, 2009) and proteomics (carpel proteomics; Baerenfaller *et al.*, 2008). Interestingly, we found that neither genes encoding for phosphoproteins from the unpollinated gynoecia and gynoecia 1hap showed signs of differential expression in a stigma-pollination transcriptional time course, nor in a pollen tube growth time course (Supplementary table S2)

In order to assess the potential transcriptional origin of proteins represented in these two phosphoproteomes, we checked the expression of the genes encoding for the identified phosphoproteins, across different tissues (as described above) (Figure 5.5). We found that gene expression signals were low specifically in male gametophytic cell types or growing pollen tubes. This suggests that the major transcriptional source of the identified phosphoproteins is not the male gametophyte, but rather the female sporophytic tissues of the gynoecium. In line with this observation, we found that genes encoding differentially identified phosphoproteins before and after pollination showed little overlap with the previously identified genes that demonstrate a transcriptional response during pollen-stigma interactions (Figure 5.4).



**Figure 5.5:** Phosphoproteins identified from unpollinated gynoecia and gynoecia 1HAP are less expressed in male specific tissue/cells: (a) Representation of relative expression values (per gene- normalized; also termed z-scores) across the tissue atlas (Wuest *et al.*, 2010) for genes encoding phosphoproteins extracted specifically from unpollinated gynoecia. (b) Representation of relative expression values (per gene- normalized; also termed z-scores) across the tissue atlas (Wuest *et al.*, 2010) for genes encoding phosphoproteins extracted specifically from gynoecia 1HAP. (c) Phosphoproteins that are shared between the unpollinated and pollinated gynoecia tissue.

### 5.3.4 Pollination-specific cross talk: A label-free approach indicates dephosphorylation dominates phosphorylation.

By merely identifying the phosphoproteins present in the male gametophyte and stigmatic sporophytic cells, we do gain limited insights into the signaling mechanisms that are involved in the cross-talk between the reproductive partners. Hence, an analysis of phosphoproteome of

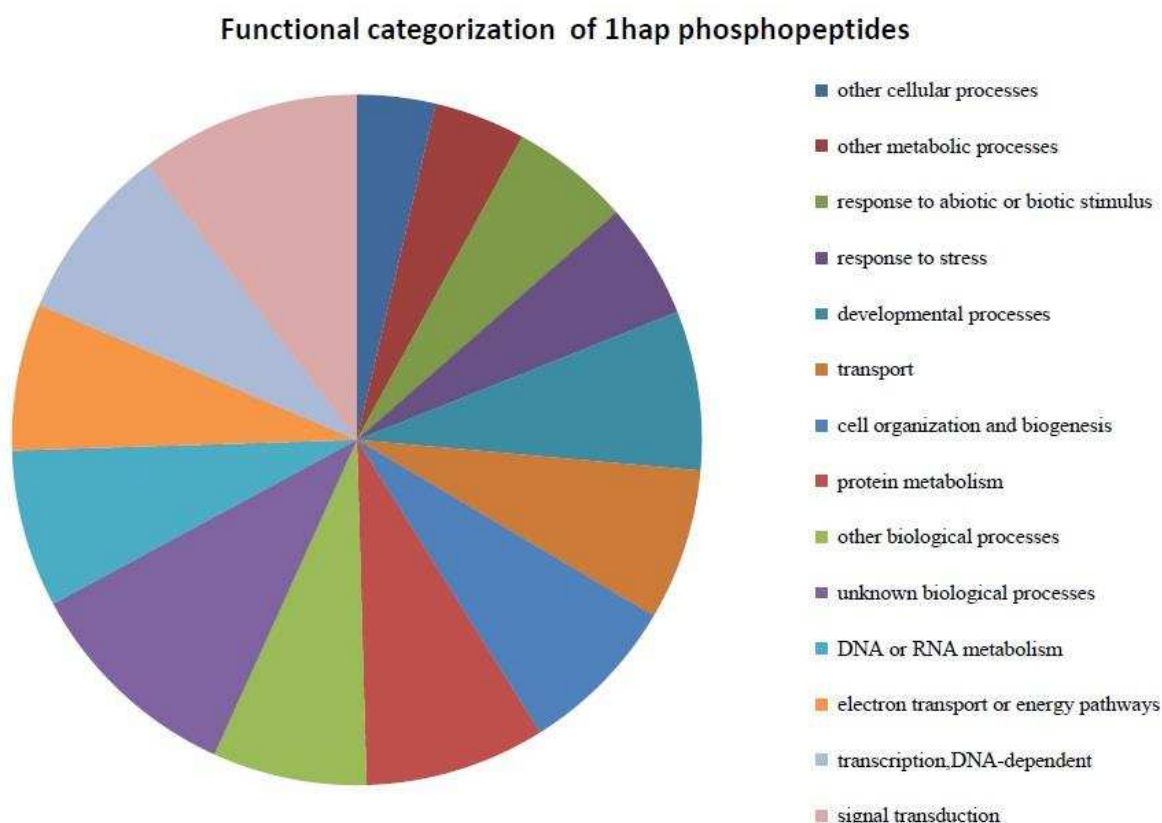


post-pollinated tissue (1HAP) can be beneficial to understand the dynamics of signaling. It is known that within one hour after pollination (Dickinson, 1995; Boavida *et al.*, 2011), pollen recognition, hydration, germination and initial pollen tube growth on stigmatic papillae takes place. These events likely rely on PTM of stored proteins, which may result in a change of the phosphorylation status (phosphorylated or de-phosphorylated), or a change in protein abundance. Therefore, a comparative label-free quantitative phosphoproteomics study between tissues 1hap and the respective unpollinated male and female partner should give insights into the potential key regulators of pollen-stigma interactions in a self-compatible pollination.

Proteins/Peptides quantification by LC-MS can be achieved either by a chemical labeling approach or by a label-free approach. The first approach uses the incorporation of stable isotopes into one or more samples and is usually done *in vivo* by introducing stable isotope-containing amino acids (SILAC) in cell culture (Ong *et al.*, 2002; Ong *et al.*, 2004) or, occasionally, *in vitro* by chemical or enzymatic means (Cagney and Emili, 2002; Ross *et al.*, 2004). The labeled peptides are heavier and are analyzed simultaneously along with the control peptides, thereby allowing them to be easily distinguished in a mass spectrometer. The labeled approach has a limitation with respect to samples sets (limited amount of samples can be analyzed in parallel) and is very expensive for routine usage. The label-free approach is more easily applied for the comparative quantification of many sample sets (no limit to samples). This type of approach is simple and cost-effective for routine usage on any sample type. The label-free approach shows high reproducibility and linearity, both at the peptide and protein levels. The limiting factor for this strategy is the sample preparation step, as it is advisable to prepare all the samples meant to be quantified in parallel, using identical reagents under similar conditions to avoid any technical variability. We used PROGENESIS LC-MS for this approach, which uses the intensity of the peptide chromatographic peak to calculate the abundance of peptides in different samples.

In order to identify the phosphoproteins and quantify the phosphopeptides for a better understanding the role of PTMs in the pollen-pistil interaction, we conducted a quantitative analysis of phosphopeptides gynoecia from post-pollinated (1HAP) using the phosphoproteome of male and female pre-pollination tissues as a reference. We focused on two major classes, referred to a potential (1) male and, (2) female regulators of pollination, respectively. Phosphopeptides which changed in phosphorylation abundance between pollen and stigma/ovary 1HAP, and are absent from the unpollinated female tissues (stigma and ovary) represent potential male regulators of pollination. Whereas, we refer to potential female

regulator of pollination to the phosphopeptides, which changed phosphorylation abundance between unpollinated and pollinated (1HAP) stigma/ovary that are absent from pollen. We detected significant changes in the abundance of 1450 phosphopeptides between pollinated and unpollinated tissues (Supplementary table S3). 76.07% (1103) of the phosphopeptides are more abundant in the reproductive partners prior to pollination. A high proportion of these phosphopeptides are from pollen (1007) and, their high abundance might be because these phosphopeptides may be highly diluted within pollinated female pistils that contain many sporophytic cells. 20.14% (292) of phosphopeptides did not change their level of abundance between the pollinated pistil and the reproductive partners prior to pollination. 3.79% (55) of the phosphopeptides showed a higher level of abundance in pollinated pistil (stigma or ovary) when compared with pollen and unpollinated pistils (stigma and ovary) (TABLE 5.3). Functional categorization on the corresponding proteins using TAIR 10 revealed categories such as “signal transduction”, “protein metabolism”, “transcription DNA dependent”, “cell organization and biogenesis”, and “unknown biological processes” (Figure 5.6). The enriched functional categories which are enriched have obvious role in pollination event especially during pollen capture and adhesion on stigmatic surface.



**Figure 5.6:** Functional categorization of phosphopeptides with a higher relative abundance in post-pollinated pistils (1HAP stigma or ovary) when compared to unpollinated pistil.

Condition Tissue Set	Higher (Pollen or Ohap Pistil)	Unchanged	Higher (1hap tissue)
<b>Pollen Vs 1hap Tissue</b> ( <i>absent</i> <i>in Ohap female tissue</i> )	<b>1007</b> 69.45%	<b>98</b> 6.76%	<b>39</b> 2.69%
<b>Ohap Vs 1hap Tissue</b> ( <i>absent</i> <i>in pollen</i> )	<b>96</b> 6.62%	<b>194</b> 13.38%	<b>16</b> 1.1%
<b>TOTAL</b>	<b>1103</b> 76.07%	<b>292</b> 20.14%	<b>55</b> 3.79%

**Table 5.3:** Label-free quantification carried out using PROGENESIS LC-MS. The table shows the percentage of phosphopeptides abundance in unpollinated and pollinated tissues. It is observed that the phosphopeptides show a high relative abundance in pollen.

Furthermore, we looked at the candidates that might play a role in pollination upon phosphorylation or dephosphorylation. These candidates changed their quantitative value (e.g. signals for a particular peptide is detected at the MS1 level and this particular peptide are then tracked across the retention time dimension. This is then used to reconstruct a chromatographic elution profile for a particular peptide after integrating all the measured value for a particular peptide; also known as feature abundance) by at least 3-fold upon pollination. To accept the presence of a feature (peptide) we set a minimum detection level of 500 units (feature abundance), below this value a feature was considered to be absent. The candidates obtained herewith were broadly classified into four categories: (1) phosphorylation candidate: putative female regulator of pollination, (2) phosphorylation candidate: putative male regulator of pollination, (3) dephosphorylation candidate: putative female regulator of pollination (4) dephosphorylation candidate: putative male regulator of pollination. (Supplement Figure S1 a, b, c) and Supplement table S3). Some of the phosphoproteins which showed high abundance in pollinated tissue belong to the calcium-binding EF hand family proteins. These proteins are involved in calcium buffering in the cytosol and in signal transduction between cellular compartments (Bentley and Réty, 2000), and as a consequence affect pollen tube growth and morphogenesis (Iwano *et al.*, 2009).

### 5.3.5 Dual phosphorylation of MPK6 at TEY motif

Among the 946 phosphorylation sites found in our study we could confirm only one Tyr-phosphosite [VTSEDFMT\*EY\*VVTR] after manually assigning fragment ions in the spectra to the expected ion type. Furthermore, Tyr-phosphorylation was confirmed by using an additional search engine, X-tandem (Fenyő and Beavis, 2003), with a  $-\log_{10}$  expectation value of 4.54 (FIGURE 5.7a). Additionally, Peptide classifier analysis confirmed that this peptide is a class 1a



2009). MAPKs are further classified into four sub-classes (A-D), based on their sequence similarity (Rodriguez *et al.*, 2010). Classes A-C contain MPKs with a specific TEY amino acid motif a typical example of which is MPK6 belonging to class A. Whereas, class D (MPK8/MPK15) have a typical canonical TDY motif for which we provided experimental evidence in our previous study (Mayank *et al.*, 2012). Group A MPKs, *i.e.*, MPK6 and MPK3, share a high sequence similarity and phosphorylate the TEY motif even then our Tyr-phosphorylated peptide [VTSE<sup>S</sup>DFMT\*EY\*VVTR] is conserved only in MPK6. In MPK3 [TSE<sup>N</sup>DFMTEYVVTR] the serine (S) (marked in red) is replaced by asparagine (N) in the corresponding sequence. Although, it has been known that dual TEY phosphorylation is important for the activation of class A MAPK (Ichimura *et al.*, 2002) there is no definite experimental evidence to date. Here, we report the dual phosphorylation of MPK6 TEY motif. MPK6 is known to play various roles in *Arabidopsis*, it is involved in cold stress, ethylene synthesis, stomata development, camelexin synthesis, pathogen, and JA signaling. MPK3 and MPK6 together have an important role in normal anther lobe formation, anther cell differentiation (Hord *et al.*, 2008). They also play a crucial role in ovule development, mutant display severe abnormalities in integument development with arrested cell divisions at later stages of ovule development (Wang *et al.*, 2008). MPK6 has its highest expression in mature pollen compared to other *Arabidopsis* tissues. MPK6 expression is highest in pollen, increasing consistently from uninucleate microspores (UNM) to mature pollen grains (MPG) (FIGURE 5.7 b, c). A similar observation was made in our label-free quantitative analysis where we see a higher abundance of phosphopeptides in pollen; however, non-phosphopeptides showed relatively higher expression in the pre- and post-pollinated female tissue. Our results show that MAPK signaling cascade is active during pollen-pistil interactions and may play a key role in the pollen development processes, upon pollination.

## 5.4 Experimental Procedures

### 5.4.1 Plant material and sample collection

The study was performed using *Arabidopsis thaliana* (L.) Heyhn (accession Columbia) *wild-type* and mutant *delayed-dehiscence2-2* (*dde2-2*) grown in the greenhouse (22°C during the day and 16°C during the night, 16 h light/8 h dark, humidity 60%) for 4 weeks. Pollen collection was carried out as described previously by Mayank *et al.*, 2012. Stigma and pistil were collected separately from *dde2-2* gynoecium for unpollinated (UP) and 1HAP. 1HAP gynocia was pollinated with wild-type *Col* plants. All the tissues were collected on liquid nitrogen and immediately stored at -80°C for further analysis. A control check was performed using aniline

blue staining for comparing the growth of pollen tube in both ecotypes. Two collections were performed to generate biological replicates.

#### 5.4.2 Protein extraction

Protein extraction was performed sequentially in two steps. Before the extraction tissue was homogenized in liquid nitrogen using a Potter homogeniser (FisherScientific, <http://www.fishersci.com/>). Firstly, soluble proteins were extracted using 100 mM HEPES/KOH pH 7.5, 15 mM EDTA, 25 mM NaF, 0.5% polyvinylpyrrolidone (PVP), 50 mM sodium pyrophosphate (NaPP) and 5% glycerol, plus an inhibitor cocktail for phosphatases (caliculin A, K252a) and proteases (Roche; <http://www.roche.com/>) followed by membrane proteins extraction using 100 and 50 mM triethylammonium bicarbonate (TEAB). All extractions were performed using an **Airfuge**<sup>®</sup> (Beckman Coulter, <https://www.beckmancoulter.com>), and the runs were performed at 117-124 kPa for 45 mins. Soluble proteins were precipitated by adding four volumes of ice-cold acetone and incubating for 2 h at -20°C. Membrane proteins were precipitated using methanol/chloroform extraction (Wessel and Flügge, 1984). Protein pellets were briefly dried at 37°C and resuspended in 50 mM TEAB buffer. Detailed protocol can be obtained from (Mayank *et al.*, 2012).

#### 5.4.3 Phosphopeptide enrichment

*Strong cation exchange:* Digested peptides were desalted using Sep-Pak reverse-phase cartridges (Waters, <http://www.waters.com/>) and pre-fractionated using strong cation exchange (SCX). Peptides were dissolved in buffer A (7 mM KH<sub>2</sub>PO<sub>4</sub>, pH 2.65, in 30% acetonitrile), and loaded onto a 2.1 ×200 mm polysulfoethyl aspartamide A column (PolyLC, <http://www.polylc.com/>) on an Agilent HP1100 binary HPLC system (Agilent Technologies, <http://www.home.agilent.com/>). Peptides were eluted in fractions with an increasing KCl gradient (10–40 min, 0–30% buffer B; 40–60 min 30–100% buffer B; buffer B consisted of 7 mM KH<sub>2</sub>PO<sub>4</sub>, pH 2.65, and 350 mM KCl in 30% acetonitrile). Eluted peptides were pooled into six fractions based on the chromatogram, and desalted again using Sep-Pak reverse-phase cartridges.

*IMAC* .Each fraction was then enriched using immobilized metal affinity chromatography (IMAC) as previously described by Mayank *et al.*, 2012. Enriched peptides were added to spin column along with Phos-select resins (40-60 µl; 50% slurry) (Sigma, <http://www.sigma-aldrich.com/>). Phos-select protocol was performed. Eluted peptides were dried and stored at -20°C.

#### 5.4.4 Data acquisition, peptide identification and label-free quantification

Dried peptides were re-suspended in 3% acetonitrile and 0.1% formic acid, and analyzed on LTQ Orbitrap XL mass spectrometer (Thermo Scientific, <http://www.thermoscientific.com>).

For each SCX fraction, a separate ProgenesisLCMS analysis was performed. In the end, we used “combine multiple fractions” within Progenesis to have the global quantitation. Individual SCX fraction Thermo raw files were loaded in ProgenesisLCMS software (Version 3.0, Nonlinear) for label free quantification. After selecting the sample with the maximum features as aligning reference 3-5 seeding vectors were provided for each file followed by automatic alignment provided by the software. Only the gradient relevant part of the features was kept while the loading phase in the first 16 minutes and the wash at the end of the LCMS run as well as all features with one charge or more than four charges were masked and excluded from further analysis. For each feature a maximum of four MS/MS spectra are exported using de-isotoping and de-noising option and only exporting the top 200 peaks per spectrum into a Mascot generic file (mgf). The mgf files are searched with Mascot server 2.3 using *Arabidopsis* TAIR 10 database containing various contaminants (e.g. human keratin, bovine proteins) (Berndt *et al.*, 1999). Search parameters used were: 10ppm peptide mass tolerance and 0.8 Da fragment mass tolerance, one missed cleavage allowed by trypsin, carbamidomethylation as fixed modification and methionine oxidation along with phosphorylation of serine, threonine and tyrosine as variable modifications. The results (only rank one and peptides scoring above 20 ion score) are re-imported into Progenesis LC-MS.

After the alignment and feature exclusion and identification with Mascot, samples were classified into the appropriate experimental group (Pollen, unpollinated stigma, unpollinated pistil, 1hap stigma and 1hap pistil) and raw abundance of all features was normalized against the five features of the spiked in fetuin digest. Normalization corrects the factors resulting from technical variability during experimental setup. Once the sample set is normalized it is then followed by statistical analysis using ANOVA (Arcsinh transformed normalized abundances for one-way analysis of variance

For quantification, all peptides with Mascot score >20 and  $P < 0.01$ , were included for further analysis. The peptide measurements of the combined-fraction analysis were exported into a CSV file and all the analysis shown in the paper was conducted with the help of Microsoft excel.

## 5.5 References

- Alexa, A., Rahnenführer, J. and Lengauer, T.,** (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics (Oxford, England)*, 22(13), 1600–7.
- Anita Lewit-Bentley\* and Stéphane Réty†,** (2000) EF-hand calcium-binding proteins. *Current Opinion in Structural Biology*, 637–643.
- Arsova, B. and Schulze, W.X.,** (2012) Current status of the plant phosphorylation site database PhosPhAt and its use as a resource for molecular plant physiology. *Frontiers in Plant Science*, 3, 132.
- Baerenfaller, K., Grossmann, J., Grobei, M.A., et al.,** (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 320(5878), 938–41.
- Baerenfaller, K., Hirsch-Hoffmann, M., Svozil, J., Hull, R., Russenberger, D., Bischof, S., Lu, Q., Gruissem, W. and Baginsky, S.,** (2011) pep2pro: a new tool for comprehensive proteome data analysis to reveal information about organ-specific proteomes in *Arabidopsis thaliana*. *Integrative Biology : Quantitative Biosciences from nano to macro*, 3(3), 225–37.
- Berndt, P., Hobohm, U. and Langen, H.,** (1999) Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis*, 20(18), 3521–6.
- Boavida, L.C., Borges, F., Becker, J.D. and Feijó, J. a,** (2011) Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in *Arabidopsis*. *Plant Physiology*, 155(4), 2066–80.
- Boavida, L.C., Borges, F., Becker, J.D. and Feijó, José A,** (2011) Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in *Arabidopsis*. *Plant Physiology*, 155(4), 2066–80.
- Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J. a and Becker, J.D.,** (2008) Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiology*, 148(2), 1168–81.



- Cagney, G. and Emili, A.,** (2002) *De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nature Biotechnology*, 20(2), 163–70.
- Chapman, L. a and Goring, D.R.,** (2010) Pollen-pistil interactions regulating successful fertilization in the Brassicaceae. *Journal of Experimental Botany*, 61(7), 1987–99.
- Dickinson, H.,** (1995) Dry stigma, water and self-incompatibility in Brassica. *Sexual Plant Reproduction*, 8(1), 1–10.
- Dickinson, H., Doughty, J., Hiscock, S.J., Elleman, C.. and Stephenson, A.G.,** (1998) Pollen-stigma interaction in Brassica, *The society of Experimental Botany* , 51, 51–7.
- Dickinson, H.. , Doughty, J., Hiscock, S J, Elleman, C.J. and Stephenson, A G.,** (1986) Pollen-stigma interactions in Brassica. *Symposia of the Society for Experimental Biology*, 80, 141–157.
- Doughty, J., Wong, H.Y., and Dickinson. H. G.,** (2000) Cysteine-rich Pollen Coat Proteins (PCPs) and their Interactions with Stigmatics (Incompatibility) and S -Related Proteins in Brassica: Putative Roles in SI and Pollination. *Annals of Botany*, 85(suppl\_1), 161–169.
- Doughty, J., Hedderson, F., McCubbin, A. and Dickinson, H.,** (1993) Interaction between a coating-borne peptide of the Brassica pollen grain and stigmatic S (self-incompatibility)-locus-specific glycoproteins. *PNAS*, 90(2), 467–71.
- Dresselhaus, T.,** (2006) Cell-cell communication during double fertilization. *Current Opinion in Plant Biology*, 9(1), 41–7.
- Durek, P., Schmidt, R., Heazlewood, J.L., Jones, A., MacLean, D., Nagel, A., Kersten, B. and Schulze, W.X.,** (2010) PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Research*, 38(Database issue), D828–34.
- Fenyő, D. and Beavis, R.C.,** (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Analytical chemistry*, 75(4), 768–74.
- Gaude.T. and McCormick.S,** (1999) Signaling in pollen-pistil interactions. *Cell & Developmental biology*, 10, 139–147.

- Grobei, M.A., Qeli, E., Brunner, E., Rehrauer, H., Zhang, R., Roschitzki, B., Basler, K., Ahrens, C.H. and Grossniklaus, U.,** (2009) Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Res*, 1786–1800.
- Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D. and Schulze, W.X.,** (2008) PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Research*, 36 (Database issue), D1015–21.
- Heslop-Harrison, Y. and Shivanna, K.R.,** (1977) The receptive surface of the angiosperm stigma. *Ann. Bot.*, 41(6), 1233–1258.
- Hirsch-Hoffmann, M., Gruissem, W. and Baerenfaller, K.,** (2012) pep2pro: the high-throughput proteomics data processing, analysis, and visualization tool. *Frontiers in Plant Science*, 3, 123.
- Hiscock S.J and Allen, A.M.,** (2008) Diverse cell signalling pathways regulate pollen-stigma interactions the search for consensus. *New Phytol*, 179(2), 286–317.
- Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J. and McCormick, S.,** (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics*, 5(18), 4864–84.
- Hony, D. and Twell, D.,** (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology*, 5(11), R85.
- Hord, C.L.H., Sun, Y.-J., Pillitteri, L.J., Torii, K.U., Wang, H., Zhang, S. and Ma, H.,** (2008) Regulation of *Arabidopsis* early anther development by the mitogen-activated protein kinases, MPK3 and MPK6, and the ERECTA and related receptor-like kinases. *Molecular Plant*, 1(4), 645–58.
- Ichimura, K., Shinozaki, K., Tena, G., et al.,** (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in Plant Science*, 7(7), 301–308.
- Iwano, M., Entani, T., Shiba, Hiroshi, et al.,** (2009) Fine-tuning of the cytoplasmic Ca<sup>2+</sup> concentration is essential for pollen tube growth. *Plant Physiology*, 150(3), 1322–34.
- Kachroo, A., Schopfer, C R, Nasrallah, M E and Nasrallah, J B,** (2001) Allele-specific receptor-ligand interactions in Brassica self-incompatibility. *Science*, 293(5536), 1824–6.

**Kandasamy, M.K., Thorsness, M.K., Rundle, S.J., Goldberg, M.L., Nasrallah, June B and Nasrallah, Mikhail E,** (1993) Ablation of papillar cell function in Brassica flowers results in the Loss of stigma receptivity to pollination. , 5(March), 263–275.

**Lord, E.M.,** (2003) Adhesion and guidance in compatible pollination. *Journal of Experimental Botany*, 54(380), 47–54.

**Malek, B. von, Graaff, E. van der, Schneitz, K. and Keller, B.,** (2002) The *Arabidopsis* male-sterile mutant *dde2-2* is defective in the ALLENE OXIDE SYNTHASE gene encoding one of the key enzymes of the jasmonic acid biosynthesis pathway. *Planta*, 216(1), 187–92.

**Mayank, P., Grossman, J., Wuest, S., Boisson-Dernier, A., Roschitzki, B., Nanni, P., Nühse, T. and Grossniklaus, U.,** (2012) Characterization of the phosphoproteome of mature *Arabidopsis* pollen. *The Plant Journal*, 1–13.

**Meimei Wang, Ying Zhang, Jian Wang, X.W. and X.G.,** (2006) A novel MAP kinase Gene in cotton (*Gossypium hirsutum* L.), GhMAPK, is Involved in response to diverse environmental stresses. *BMB Reports*, 40 (3), 325–332.

**Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M.,** (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular proteomics : MCP*, 1(5), 376–86.

**Ong, S.-E., Mittler, G. and Mann, M.,** (2004) Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. *Nature Methods*, 1(2), 119–26.

**Pina, C., Pinto, F., Feijó, J.A. and Becker, J.,** (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology*, 138(June), 744–756.

**Pitzschke, A., Schikora, A. and Hirt, H.,** (2009) MAPK cascade signalling networks in plant defence. *Current Opinion in Plant Biology*, 12(4), 421–6.

**Qeli, E. and Ahrens, C.H.,** (2010) PeptideClassifier for protein inference and targeted quantitative proteomics. *Nature Biotechnology*, 28(7), 647–50.

**Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M. a and Palanivelu, R.,** (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, 5(8), e1000621.

**Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W. and Baginsky, S.,** (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant physiology*, 150(2), 889–903.

**Rodriguez, M.C.S., Petersen, M. and Mundy, J.,** (2010) Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology*, 61, 621–49.

**Ross, P.L., Huang, Y.N., Marchese, J.N., et al.,** (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP*, 3(12), 1154–69.

**Sarker, R.H., Elleman, C.J. and Dickinson, H.G.,** (1988) Control of pollen hydration in Brassica requires continued protein synthesis, and glycosylation is necessary for intraspecific incompatibility. *PNAS*, 85(12), 4340–4.

**Schopfer, C. R.,** (1999) The male determinant of Self-Incompatibility in Brassica. *Science*, 286(5445), 1697–1700.

**Scutt, C.P., Vinauger-Douard, M., Fourquin, C., Ailhas, J., Kuno, N., Uchida, K., Gaude, T., Furuya, M. and Dumas, C.,** (2003) The identification of candidate genes for a reverse genetic analysis of development and function in the *Arabidopsis* gynoecium. *Plant Physiology*, 132(2), 653–65.

**Shi, D.-Q. and Yang, W.-C.,** (2011) Ovule development in *Arabidopsis*: progress and challenge. *Current Opinion in Plant Biology*, 14(1), 74–80.

**Shiba, H.,** (2001) A pollen coat protein, SP11/SCR, determines the pollen S-specificity in the Self-Incompatibility of Brassica Species. *Plant Physiology*, 125(4), 2095–2103.

**Swanson, R., Clark, T. and Preuss, D.,** (2005) Expression profiling of *Arabidopsis* stigma tissue identifies stigma-specific genes. *Sexual Plant Reproduction*, 18(4), 163–171.

**Wang, H., Liu, Y., Bruffett, K., Lee, J., Hause, G., Walker, J.C. and Zhang, S.,** (2008) Haplo-insufficiency of MPK3 in MPK6 mutant background uncovers a novel function of these two MAPKs in *Arabidopsis* ovule development. *The Plant cell*, 20(3), 602–13.

**Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.-J., Su, Z. and Wu, W.H.,** (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiology*, 148(3), 1201–11.

**Wessel, D. and Flügge, U.I.,** (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry*, 138(1), 141–143.

**Wuest, S.E., Vijverberg, K., Schmidt, A., et al.,** (2010) *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current biology : CB*, 20(6), 506–12.

## 5.6 Supporting Information

### 5.6.1 Supplement Figure S1 (a-d): Level plots of quantified candidates:

- (a) Phosphorylation candidate: Female regulator of pollination
- (b) Phosphorylation candidate: male regulator of pollination
- (c) Dephosphorylation candidate: Female regulator of pollination
- (d) Dephosphorylation candidate: male regulator of pollination.

NOTE: For condition (d) we only online version of the plots

### 5.6.2 Supplement Tables:

#### 5.6.2.1 Supplement table S1: Overview of Arabidopsis pollination-specific tissue phosphoproteome

The Microsoft Excel displays a table that was generated using the Pivot tables. The table lists the total number of unique phosphopeptides and corresponding phosphoproteins identified in current study individually from pollen, unpollinated female tissue (stigma and pistil), and 1hap tissue (stigma and pistil). Additionally, table list the peptide sequence with confirmed phosphorylation site (star sequence), average Mascot ion score of phosphopeptides, distribution of phosphopeptides in respective tissue set and Peptide Classifier classification of each phosphopeptides.

**NOTE:** Excel File is present in the supplementary CD

#### 5.6.2.2 Supplement table S2: Overlap of data between current phosphoproteomics and previous transcriptomics and proteomics studies on Arabidopsis pollination-specific tissue

The Microsoft Excel list shows the overlap of unpollinated and pollinated (1hap; one hour after pollination) phosphoproteome with previous transcriptomics and proteomics study. The data used for generating an overlap was extracted from transcriptomics study (0.5 hap; [Boavida et al., 2011](#)), 0.5h *in vitro* pollen tube ([Qin et al., 2009](#)) and proteomics study (carpel proteomics ([Baerenfaller et al., 2008](#))). Interestingly, we found that neither genes encoding for phosphoproteins from the unpollinated and 1hap stigma tissues showed signs of differential expression in a stigma-pollination transcriptional time course, nor in a pollen tube growth time course.

**NOTE:** Excel File is present in the supplementary CD



#### 5.6.2.3 Supplement table S3: List of label-free quantification candidates generated using PROGENESIS LC-MS (nonlinear dynamics) software

The list displays the abundance level of each phosphopeptides in respective tissue. Each phosphopeptide in the list has a Mascot score of >20 and P value of < 0.01. Based on their expression and abundance level the phosphopeptides were classified into four different classes (a) Phosphorylation candidate: Female regulator of pollination (b) Phosphorylation candidate: male regulator of pollination (c) Dephosphorylation candidate: Female regulator of pollination (d) Dephosphorylation candidate: male regulator of pollination.

**NOTE:** Excel File is present in the supplementary CD

#### 5.6.2.4 Supplementary table S4: Number of aborted and normal seed comparison between wild type and dde2-2 mutant

Statistically the seeds set between the two backgrounds are comparable.

**NOTE:** Excel File is present in the supplementary CD

## Conclusion and Outlook

---

Proteins play a vital role in living organisms as they are the major component of the physiological pathways of cells. Therefore, a global analysis and map of the proteome of an organism provides an excellent resource for future studies. After genomics and transcriptomics, proteomics gained importance in the biological sciences but owing to its higher complexity progress was slower. The genome of an organism is more or less constant at all times; however, the transcriptome and proteome differs temporally from cell to cell. Moreover, proteins are subjected to a wide variety of chemical modifications after translation through so called post translational modifications which include methylation, acetylation, glycosylation, oxidation, nitrosylation and most widely studied, phosphorylation. Phosphorylation plays a central role in cell signaling and regulates various enzymes, by either activating or deactivating them via phosphorylation and dephosphorylation.

Phosphoproteomics is a powerful approach for the global study of protein phosphorylation and the assignment of phosphorylated substrates to specific kinases. These *in vivo* and *in vitro* approaches are gaining more importance in the plant field, specifically in *Arabidopsis*, because of its role as a model system and complete sequenced genome and proteome.

In this thesis, the methodology used for conducting phosphoproteomics on *Arabidopsis* pollination-specific tissues was optimized for enrichment of phosphopeptides from a very limited amount of sample (pollen grains and stigmas), which were then applied to the dynamic changes in phosphoproteome during pollination process.

Chapter 4 of this thesis provides a detailed phosphoproteomics study of the *Arabidopsis* mature pollen. Despite its structural simplicity, pollen performs complex functions and many interactions, are necessary for the development of the pollen tube and the successful delivery of the male gametes to their female partners. Most of our knowledge about pollen development and function is based on genetics, transcriptomics and a few proteomics analyses. These are not sufficient to understand the signaling processes occurring during pollination, in particular, during pollen hydration, germination, and initial PT growth. It is known that a mature dehydrated pollen grain mostly contains all the necessary transcripts and proteins required for its initial growth. Therefore, it was important to generate a phosphoproteomics reference map containing all the potential targets of phosphorylation in the dehydrated pollen grain. This study provides a wealth of knowledge that can be related to the signaling mechanisms involved in these initial developmental processes during pollination. Some of the phosphoproteins which have been found in our study show specificity only to the pollen, confirmed by a transcriptome analysis (Chapter 3). Site-directed mutagenesis studies on identified phosphorylated residues can be used in the future to

further explore the importance of phosphorylation for pollen development processes. Furthermore, various classes of proteins, having a potential role in PT growth were found to be phosphorylated, providing evidence for post-translational regulation.

In Chapter 5, the phosphoproteome map of pollination-specific tissue of *Arabidopsis* has been presented, where I compared the differential phosphorylation patterns of unpollinated and pollinated pistils. Interestingly, 1hour after pollination (hap) tissue showed an over-representation of GO categories with an obvious role in pollination such as “protein auto-phosphorylation,” “pollen-pistil interaction,” “positive regulation of cell size,” “arginine biosynthetic process,” and “translational initiation”. Many of these categories have an obvious role in pollen-stigma interaction. In the later section of the chapter we describe label-free approach for checking the abundance level of some of the differentially regulated phosphopeptides. These peptides act as male or female regulator of pollination.

Finally the data generated in this study also acts as a reference map for comparison of the downstream signaling events of pollination in self-compatible species with self-incompatible species (*Arabidopsis lyrata*).

## Acknowledgments

---

To begin with I would like to express my profound gratitude to my supervisor Prof. Ueli Grossniklaus. He provided me with an excellent opportunity to conduct novel phosphoproteomics study on *Arabidopsis thaliana* reproductive specific tissue and kept motivating and encouraging me during all the phases of my doctoral studies. I am extremely grateful for all the advice and suggestion I received from him which gave this project a good direction and helped me to publish my results in reputed international journal.

I would like to extend my gratitude to Prof. Christian von Mering and Dr. Thomas Nuhse for their co-supervision. Special thanks go to Dr. Thomas Nuhse for all his help and support, without which this Ph.D. would not have been possible.

A great deal of thanks goes to Center of Model Organism Proteome (C-MOP) and Functional Genomics Center Zurich (FGCZ) proteome and bioinformatics facility for their fruitful discussion and guidance. Special thanks to Ralph Schlapbach for providing access to facility of FGCZ. Dr. Bernd Roschitzki and Dr. Paolo Nanni for their support with Mass-Spectrometry analysis, Claudia Fortes for her support with HPLC and finally Dr. Jonas Grossmann for his tremendous support, bioinformatics analysis and immediate supervision related to data analysis.

I would like to thank the University Research Priority Program (URPP) Systems Biology/ Functional Genomics for funding.

I thank Grossniklaus's present and ex- lab members for the useful discussion and very friendly atmosphere. Special thanks to Dr Valeria Gagliardini, Arturo Bolanos, Peter Kopf and Daniela Guthorl for their technical help and Daniela Guthorl for regular assistance with German translations. Andrea Monica Schauer, Dr. Aurelien Boisson-dernier, Dr. Christian Heichinger and Dr. Stephen Schauer for their continuous help and support related to the progress of my project. Special thanks to Dr. Aurelien Boisson-Dernier for reading my manuscripts and providing his expert comment which helped to give a better shape to the manuscripts. Kostas, Michael, Arco, Milka, Wendelin, Arturo, Christian, Daniela, Joana, Aurelien, Hannes, Afif, Lena, Christian, Guillaume, Stephan, Rita, Nuno, Marek and all the members of Grossniklaus group for their friendship and the nice time I spent with you all during my Ph.D. I also would like to thank Dr. Hanspeter Schob, Monica Frani, Dr. Greco Kroes, and Jasmin Unternaher for their administrative support.

Finally I am thankful to God for such loving and caring grandparents, parents, Sister (Maiteryi Pathak), Niece (Gungun), partner (Nidhi) and my sher (dog) Jaisel. Without their love and continuous support in my daily life it is very hard to imagine life.

Lastly I would like to thank my entire social circle around the globe for their help, support, and leisure activities which motivated me during my stressful days.

### **8.1 Supplement Figure S1 (a-d): Level plots of quantified candidates:**

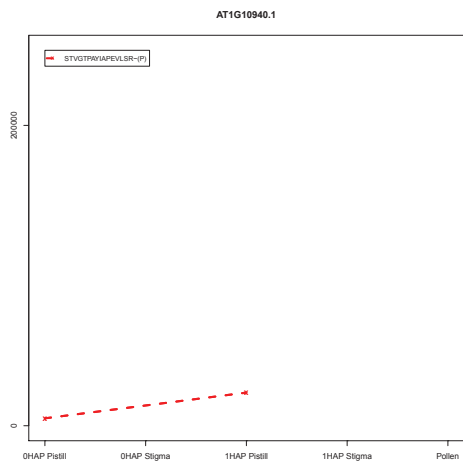
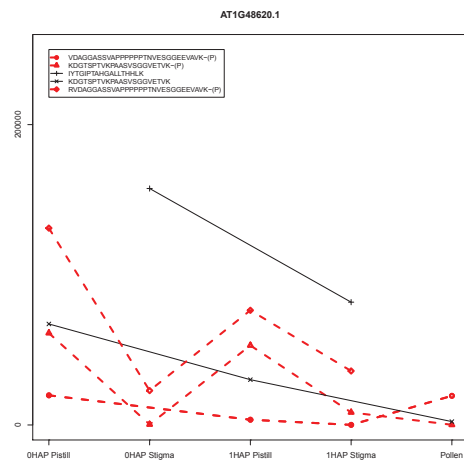
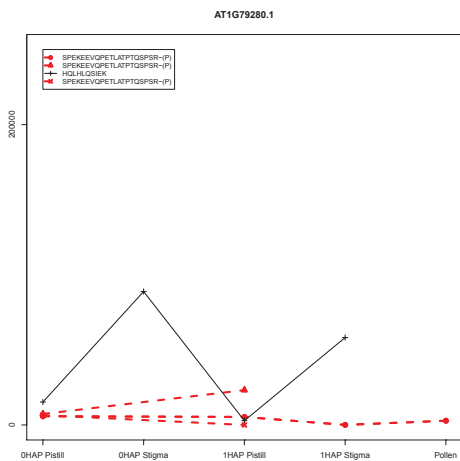
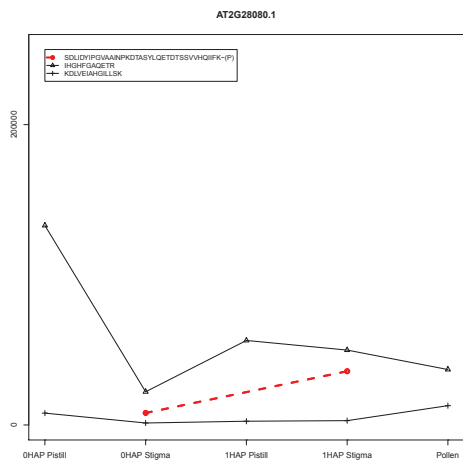
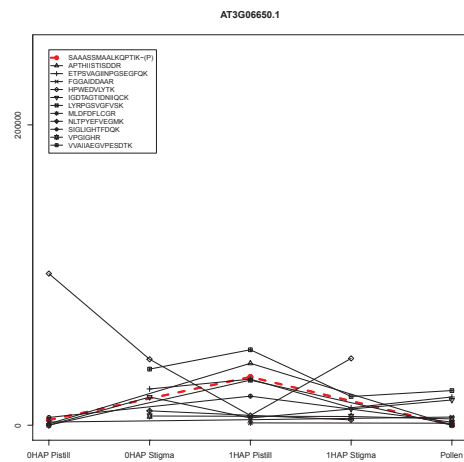
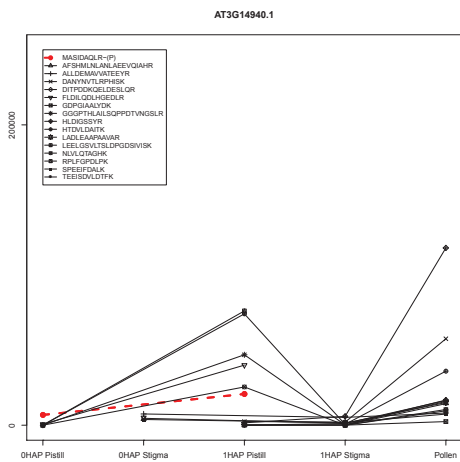
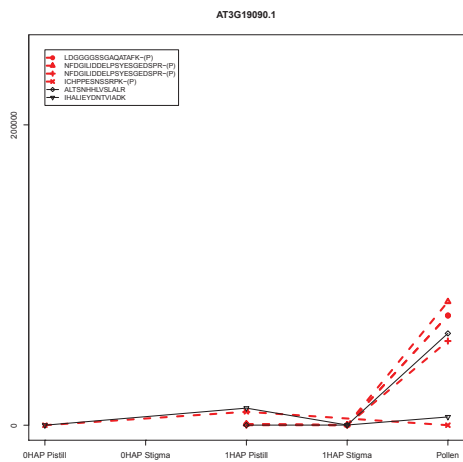
### **8.2 Supplement Tables: Presented in the CD along with the thesis.**

NOTE: All the excel file are present in the CD supplement with the thesis.

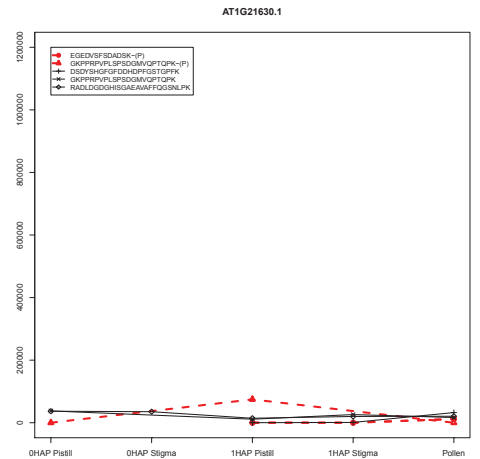
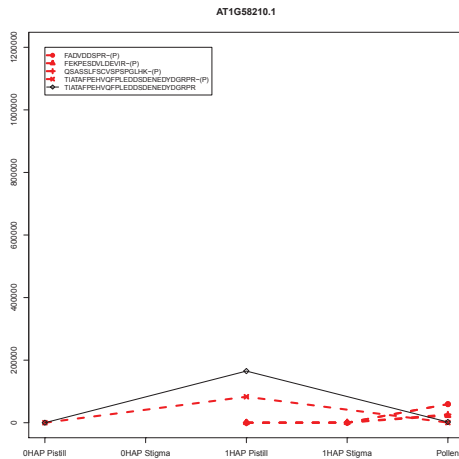
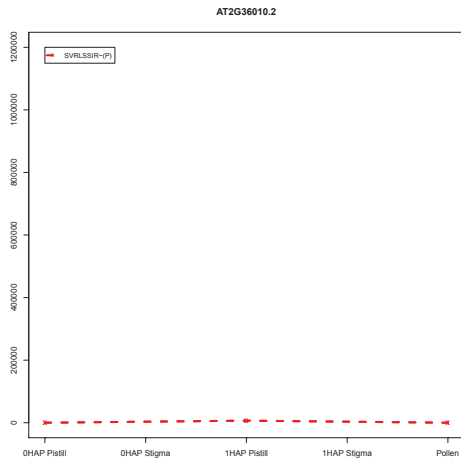
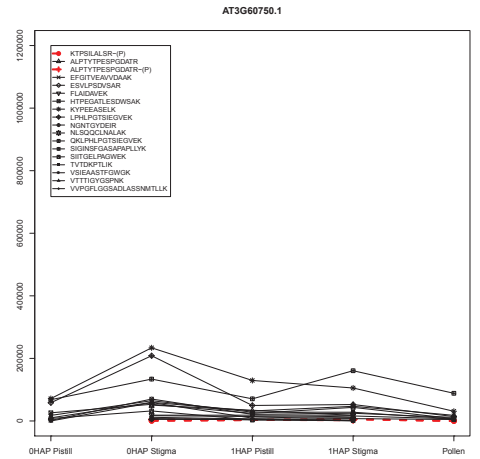
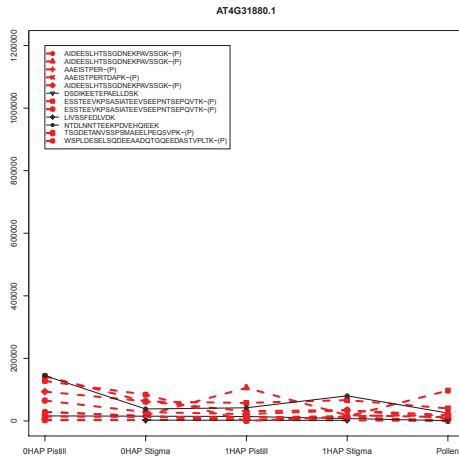
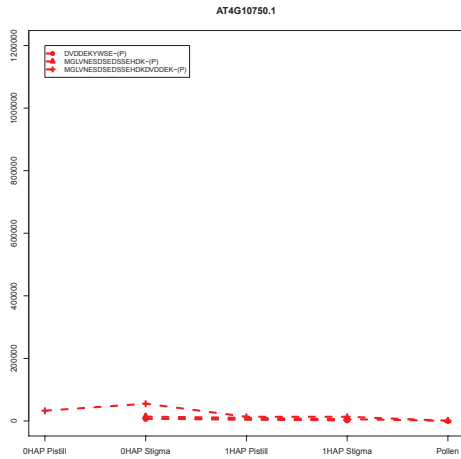
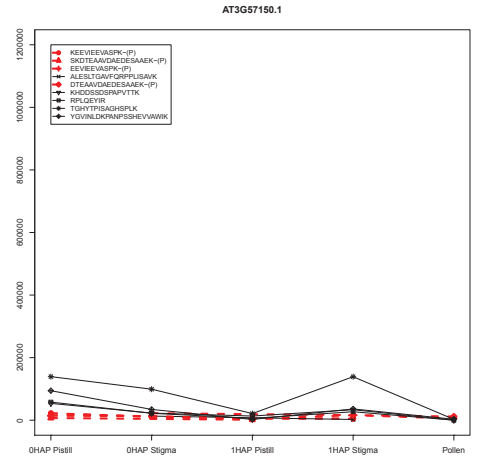
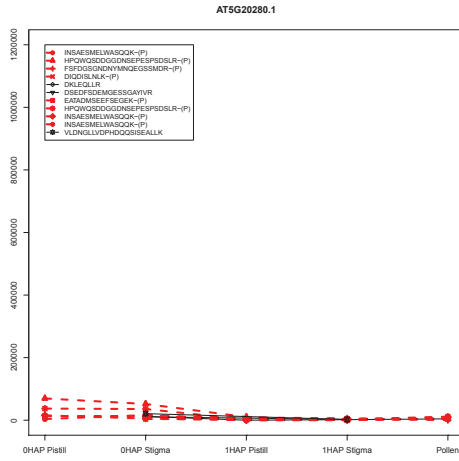
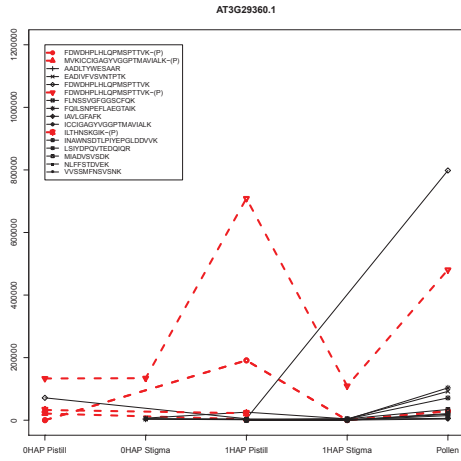


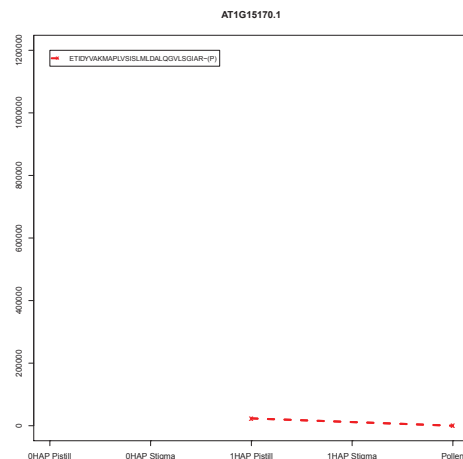
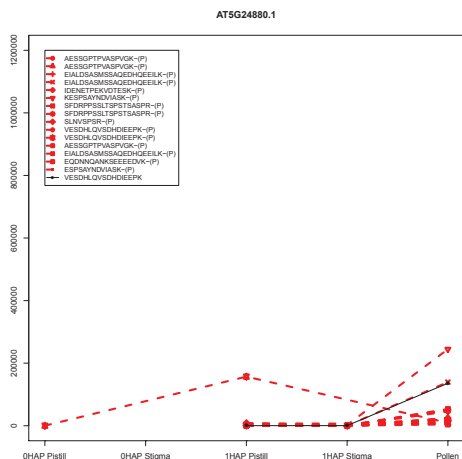
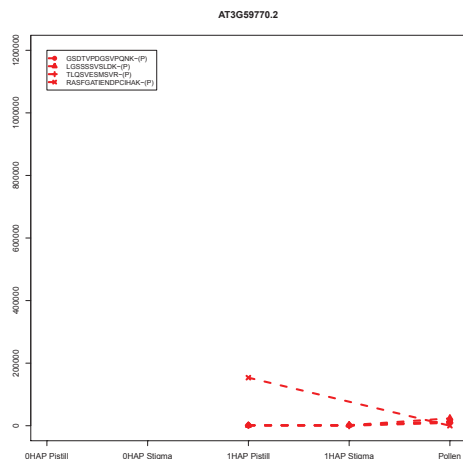
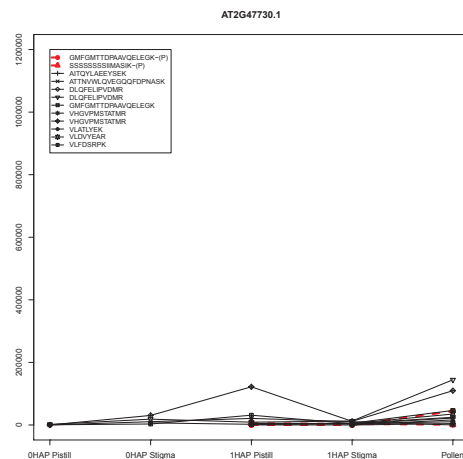
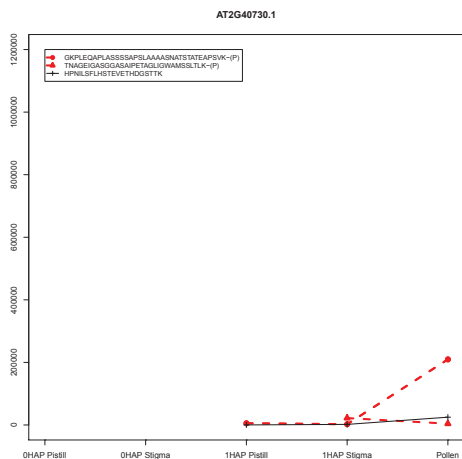
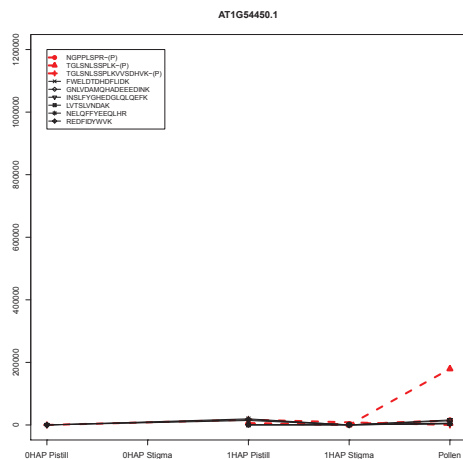
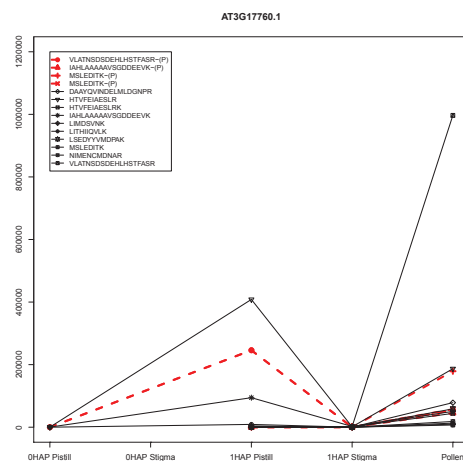
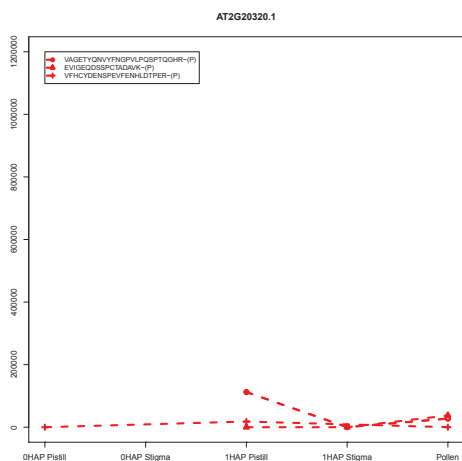
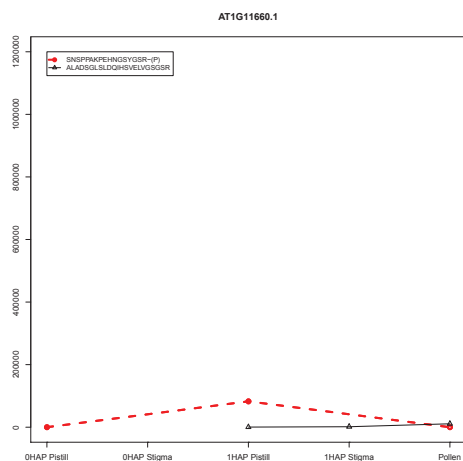
Supplement Figure S1 a: Phosphorylation candidate: Female regulator of pollination

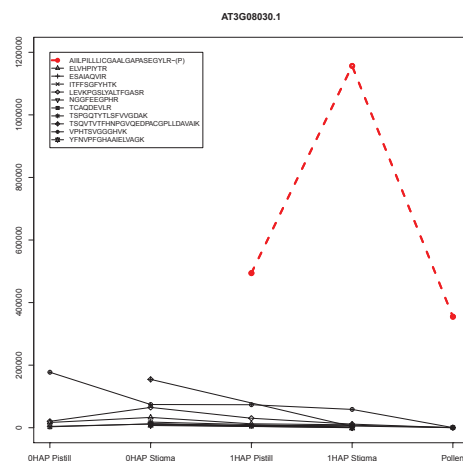
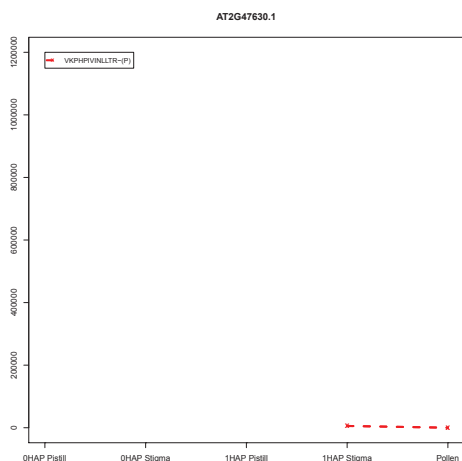
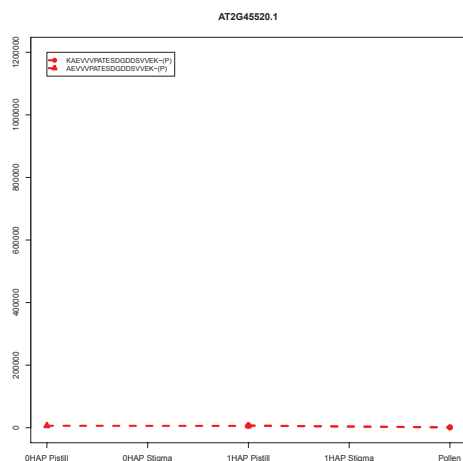
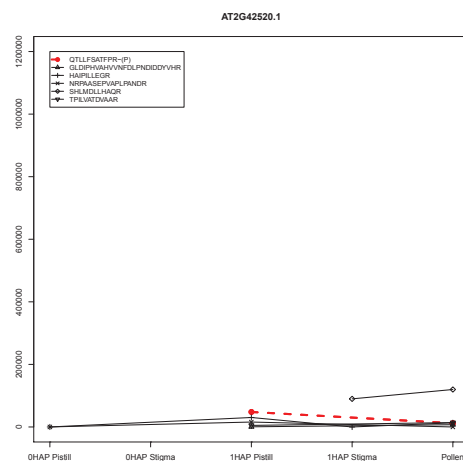
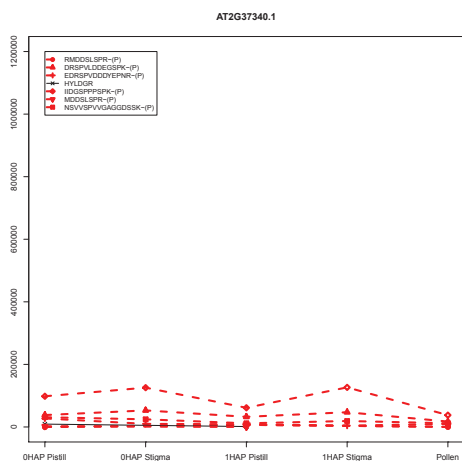
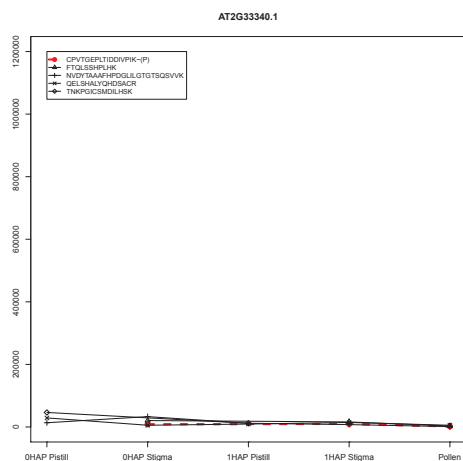
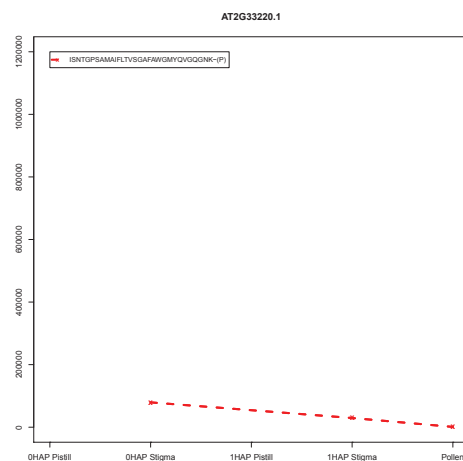
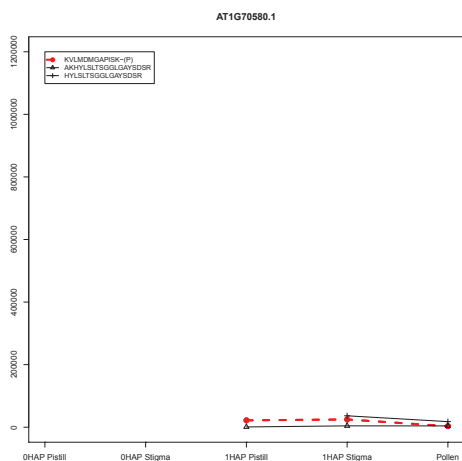
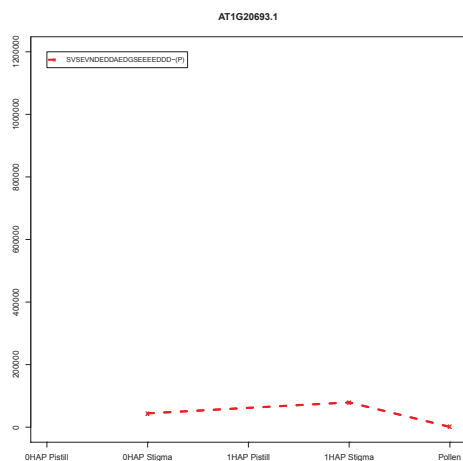


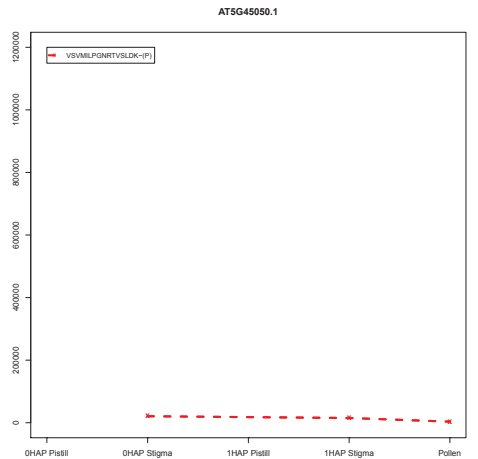
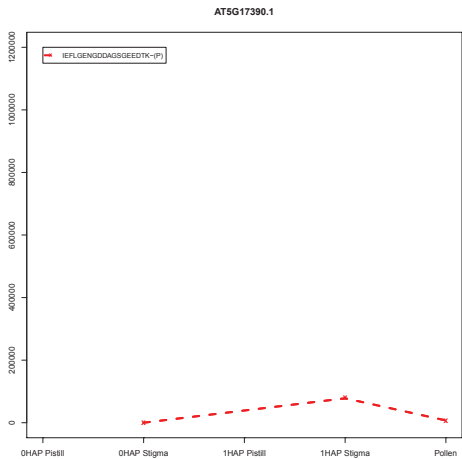
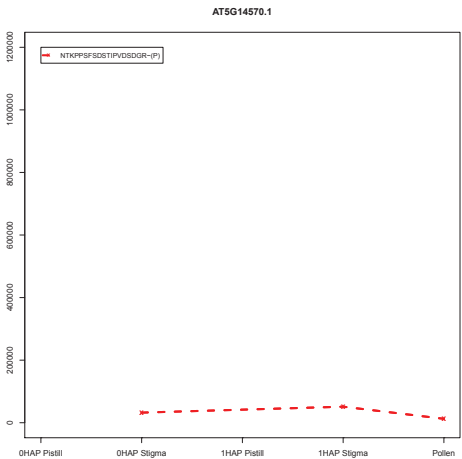
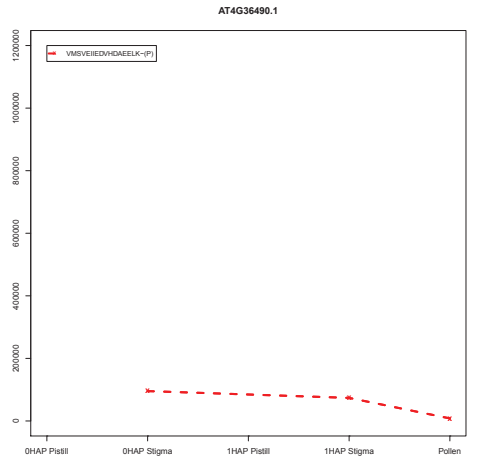
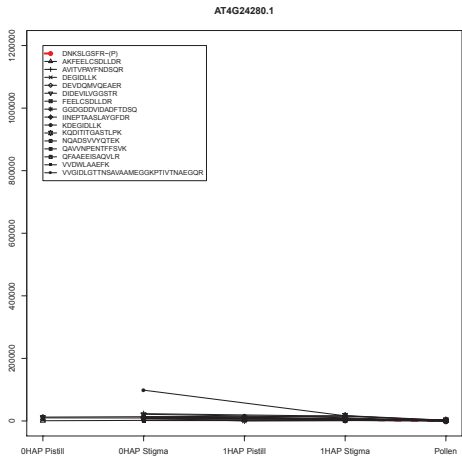
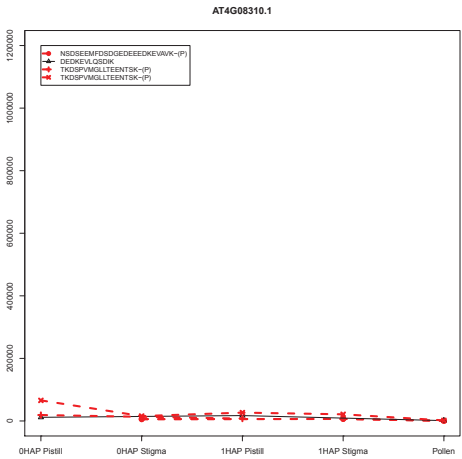
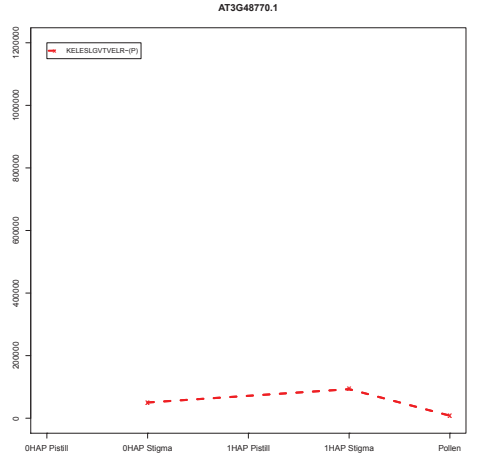
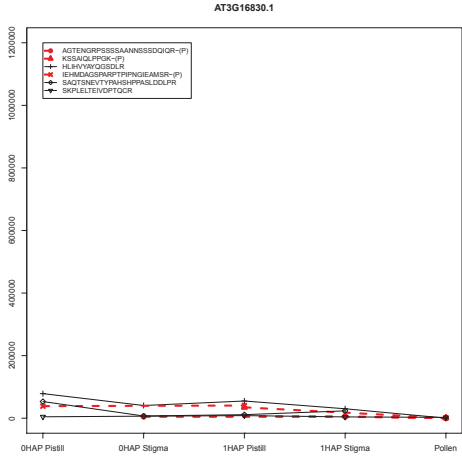
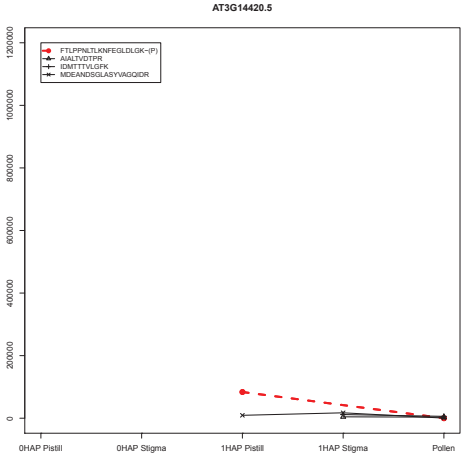


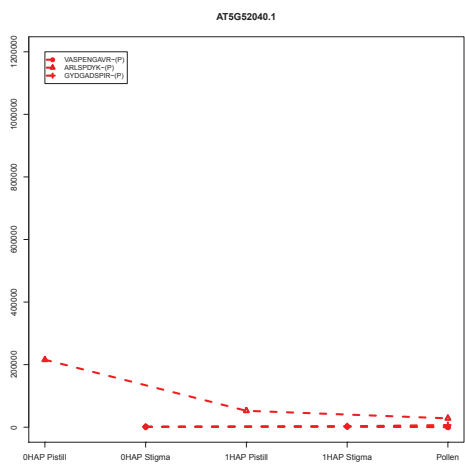
# Supplementary figure S1 b: Phosphorylation candidate: male regulator of pollination







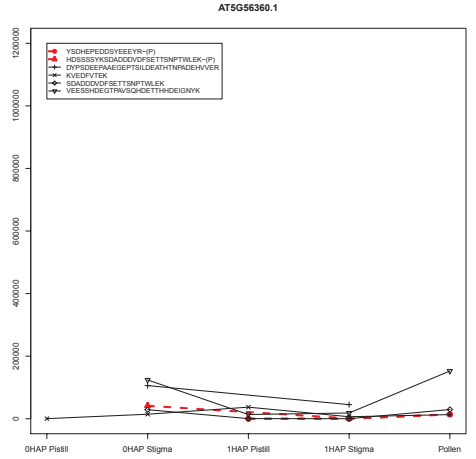
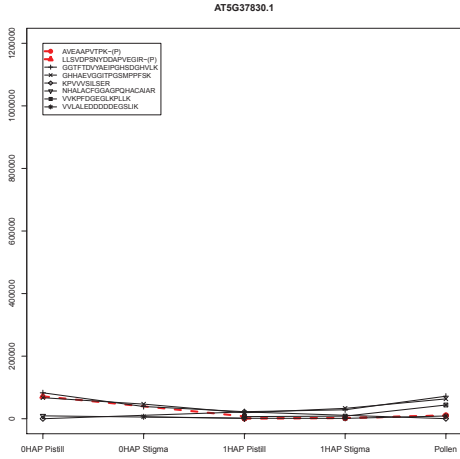
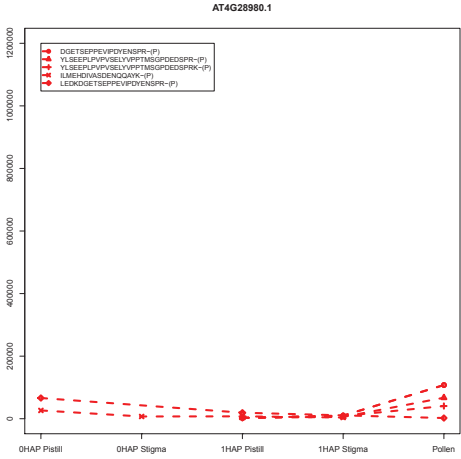
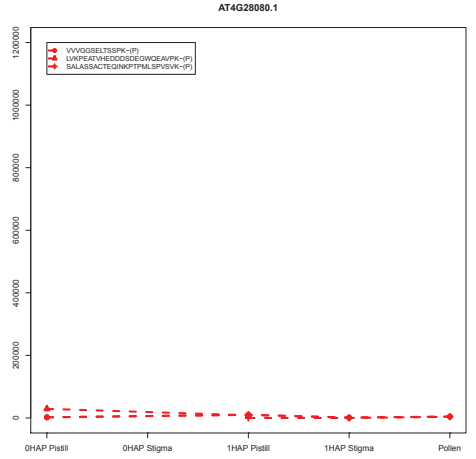
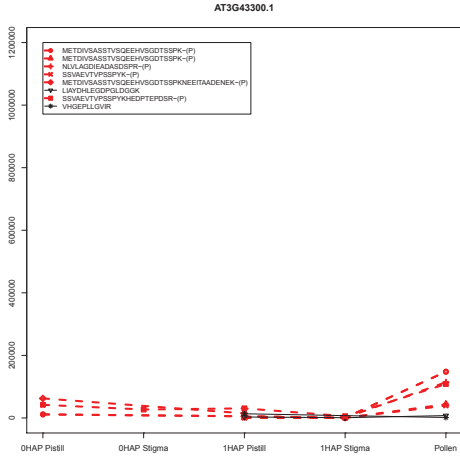
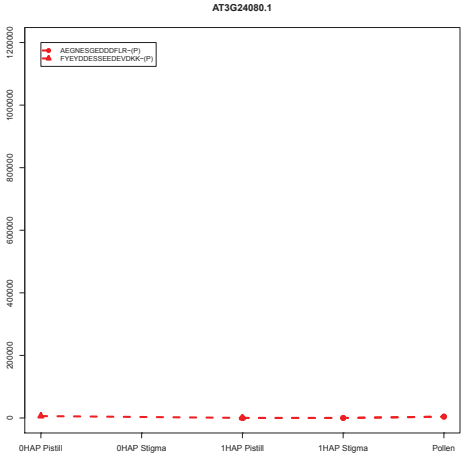
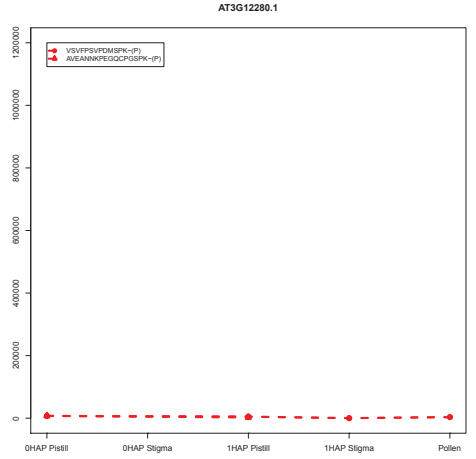
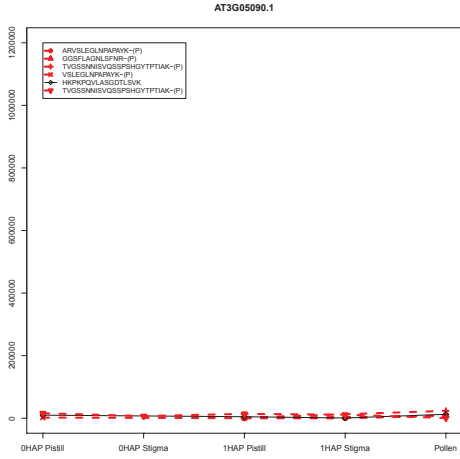
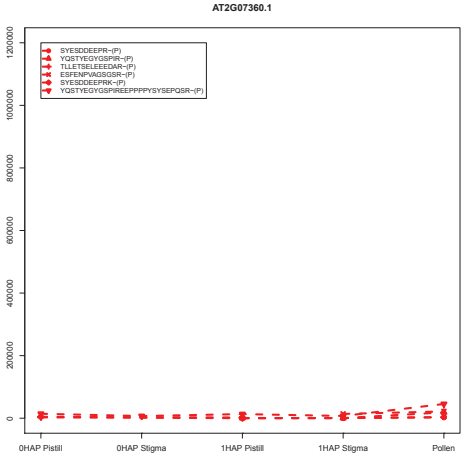


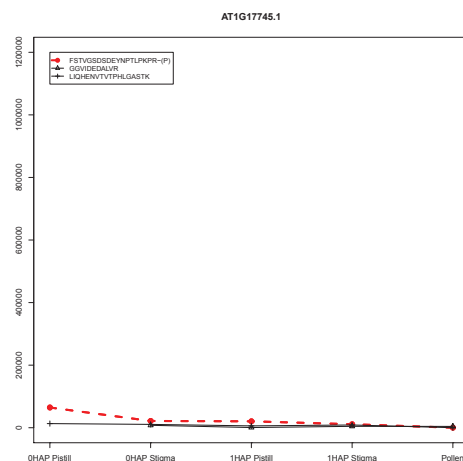
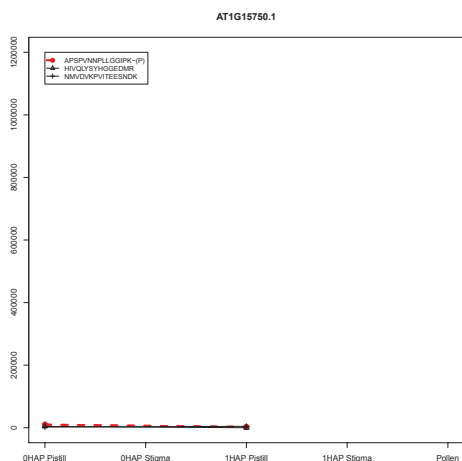
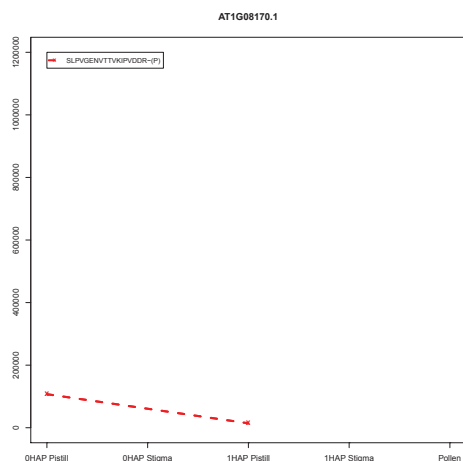
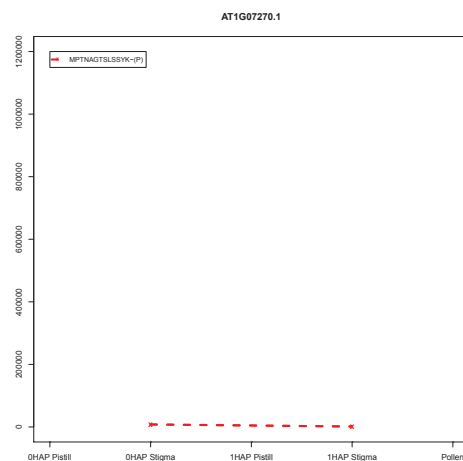
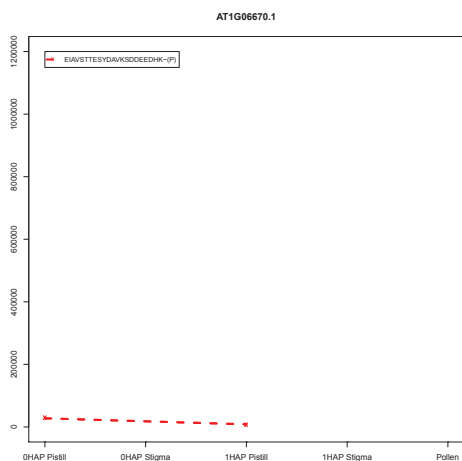
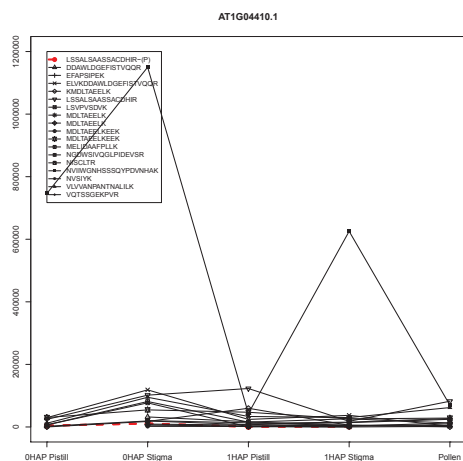
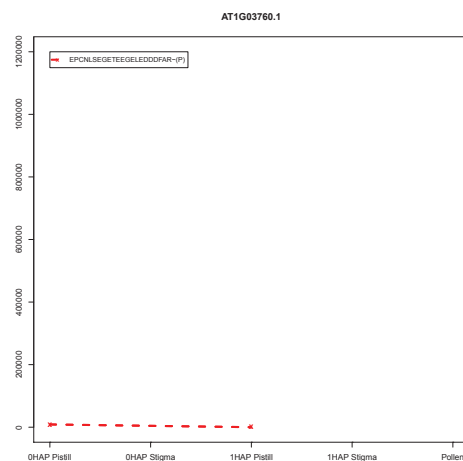
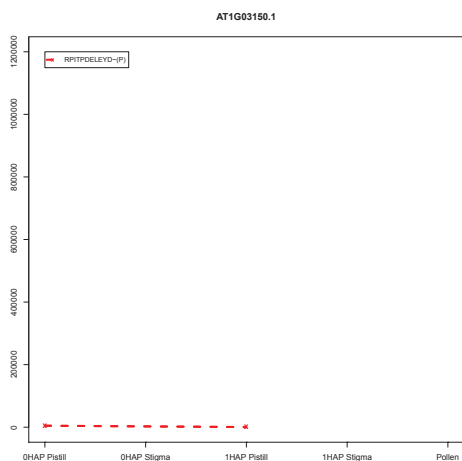
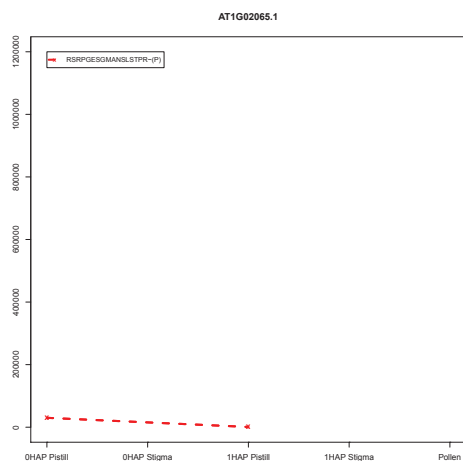


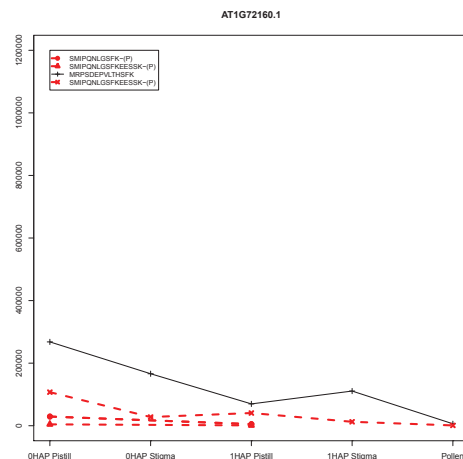
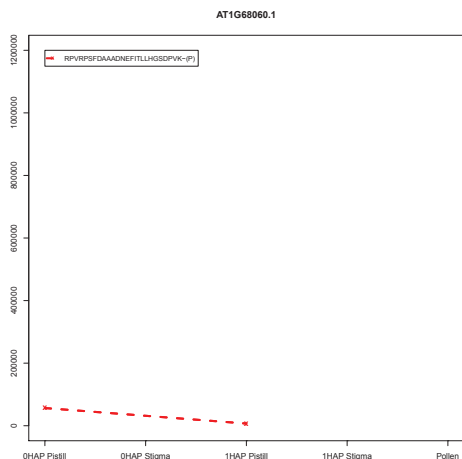
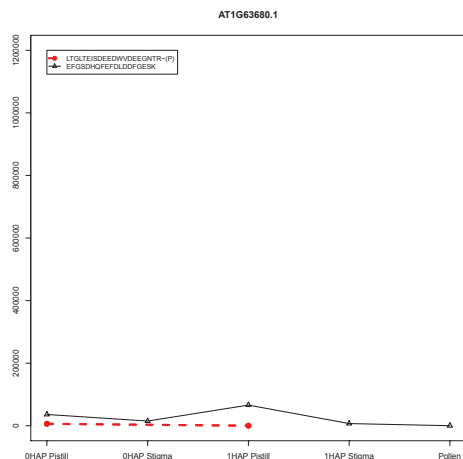
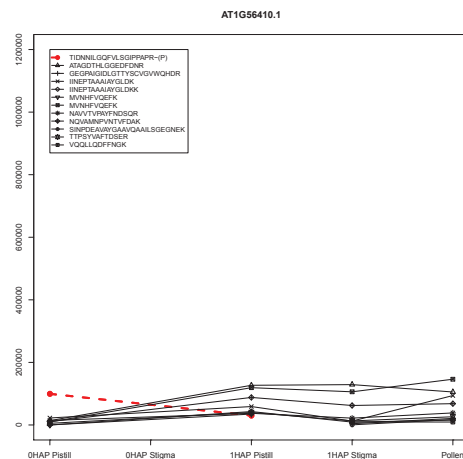
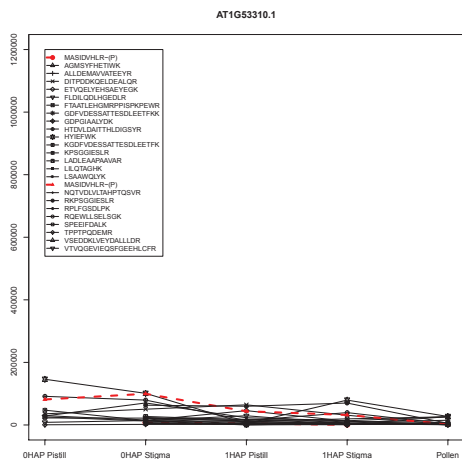
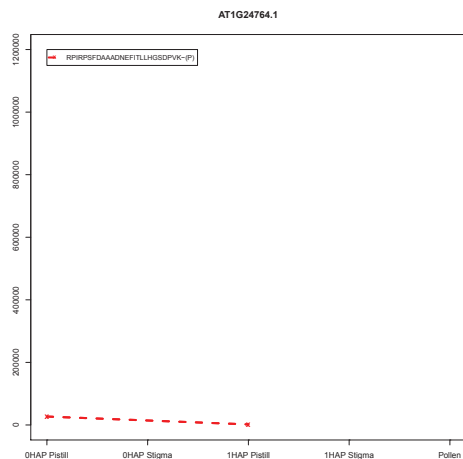
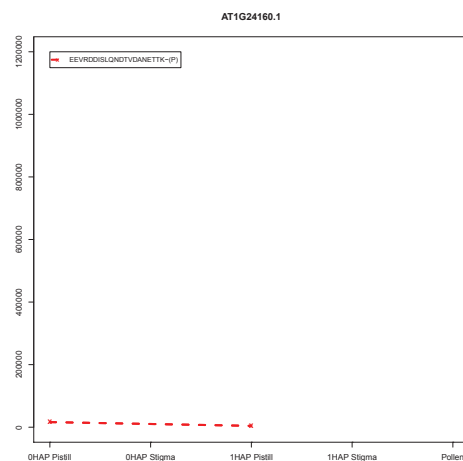
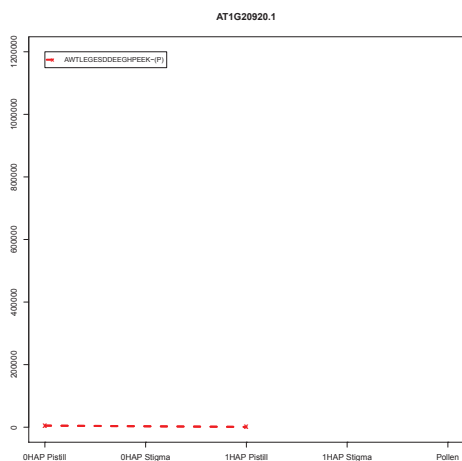
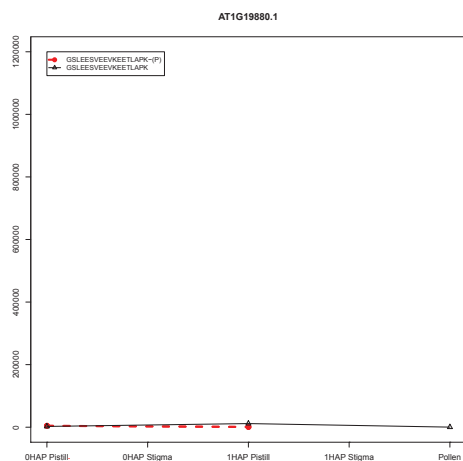


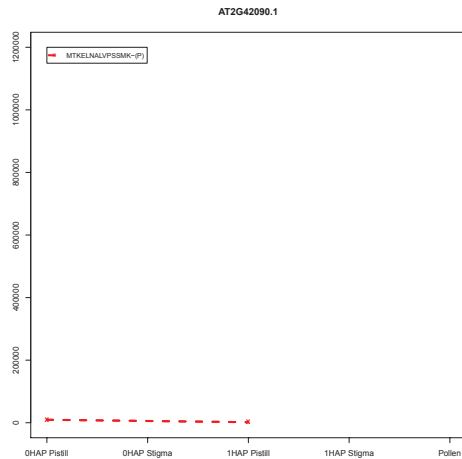
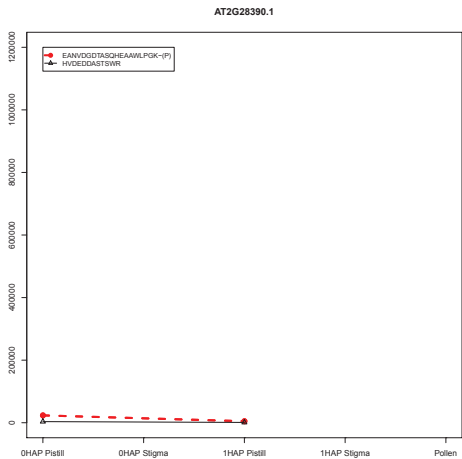
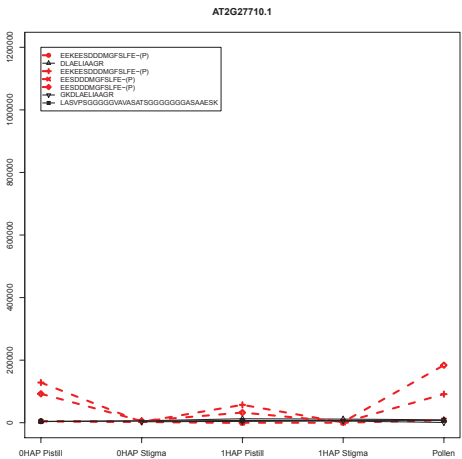
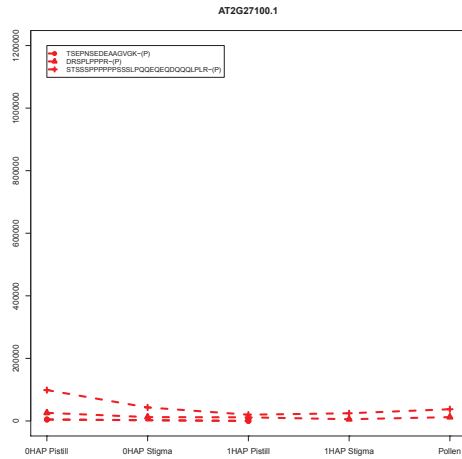
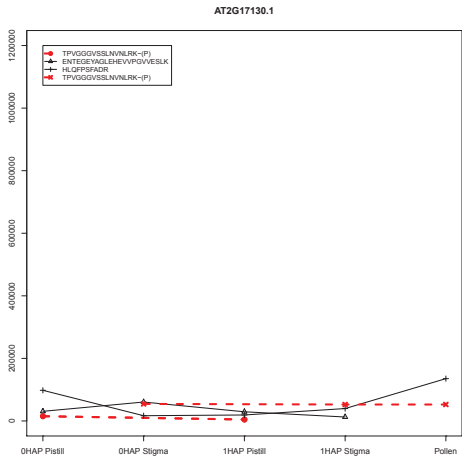
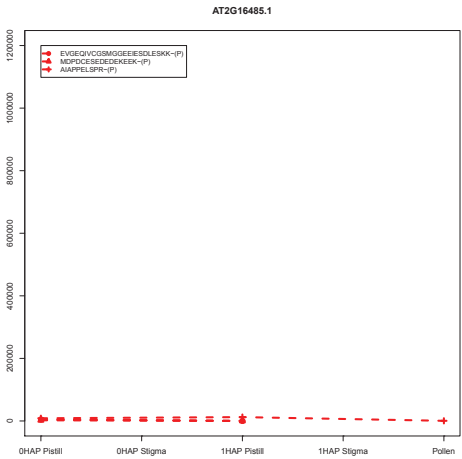
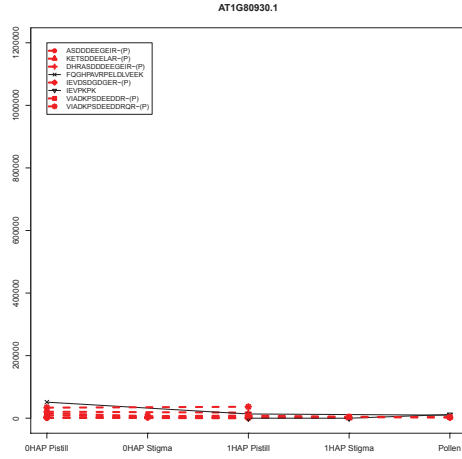
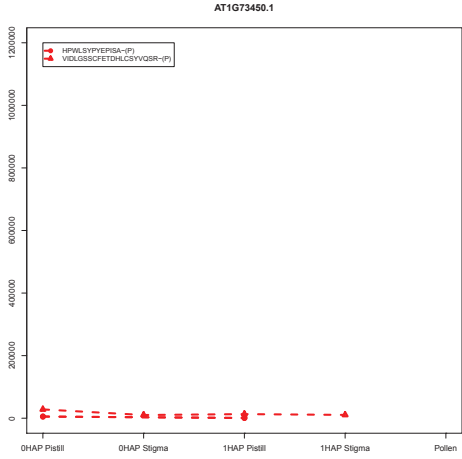
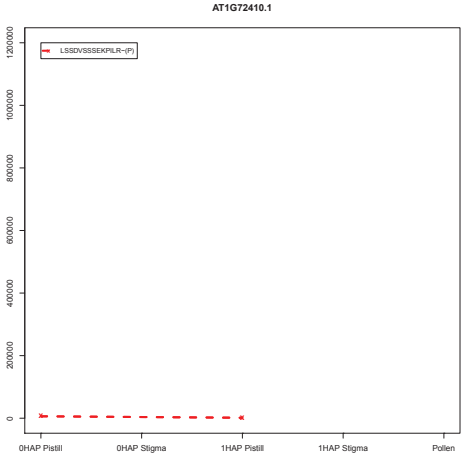
# Supplementary figure S1 c: Dephosphorylation candidate: Female regulator of pollination

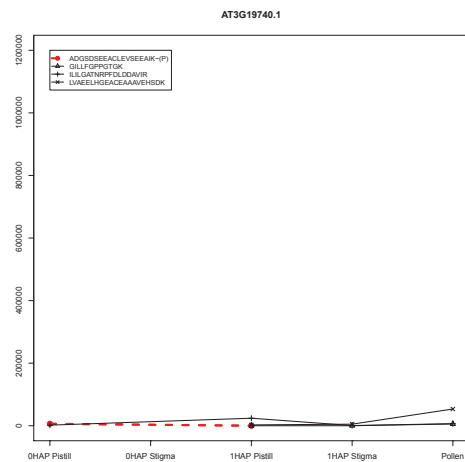
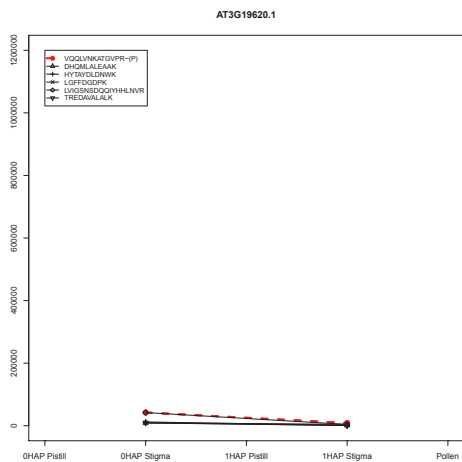
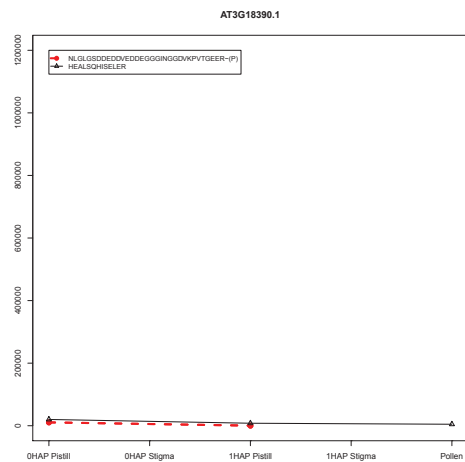
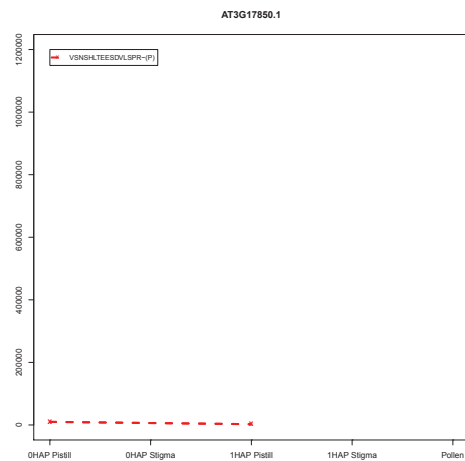
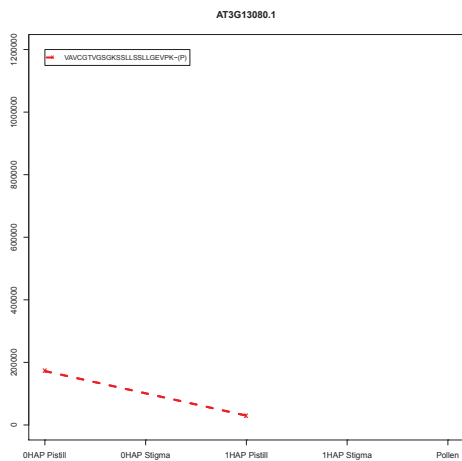
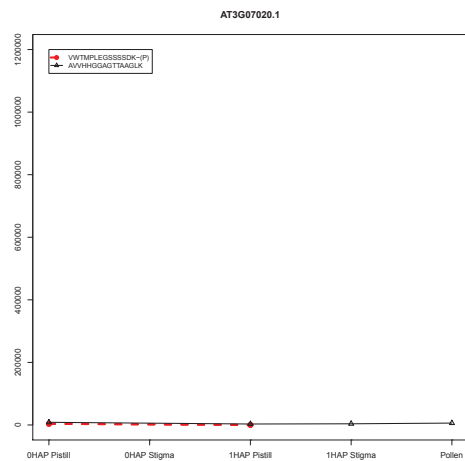
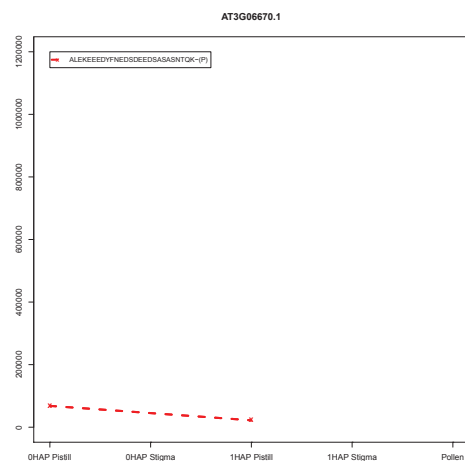
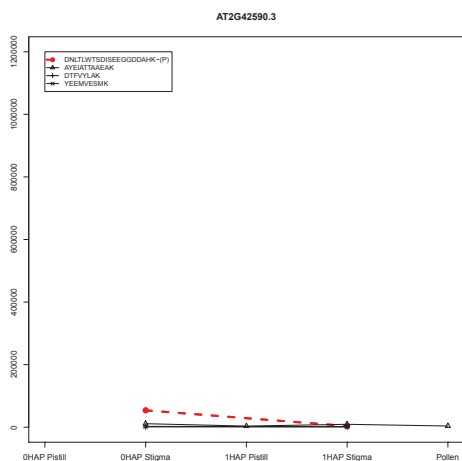
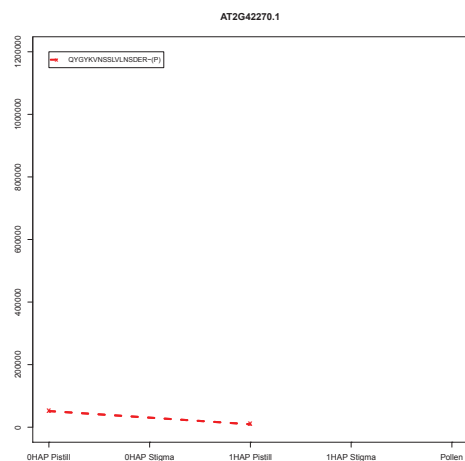


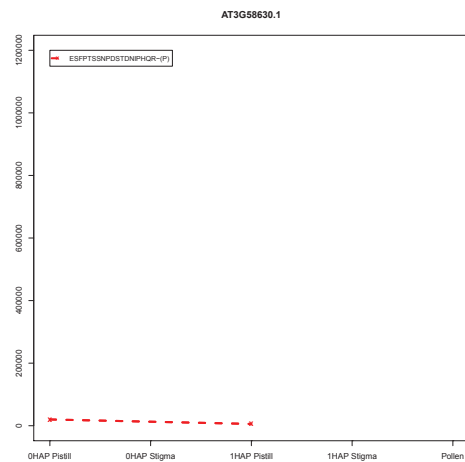
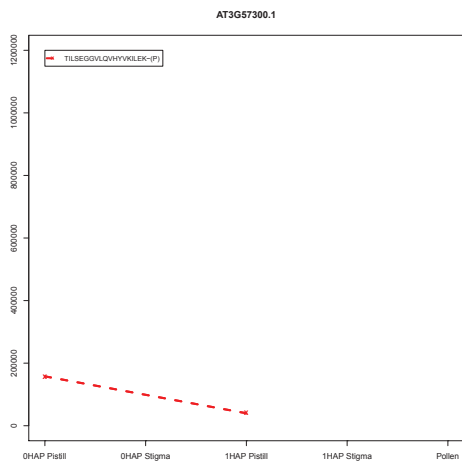
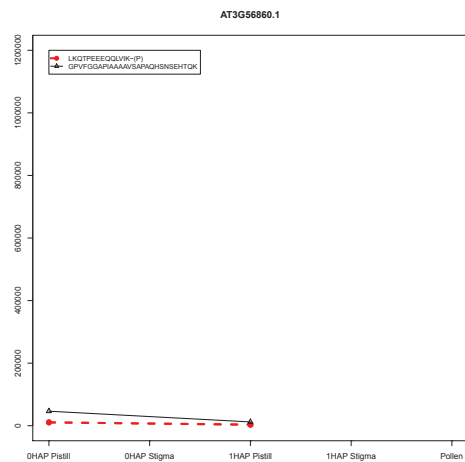
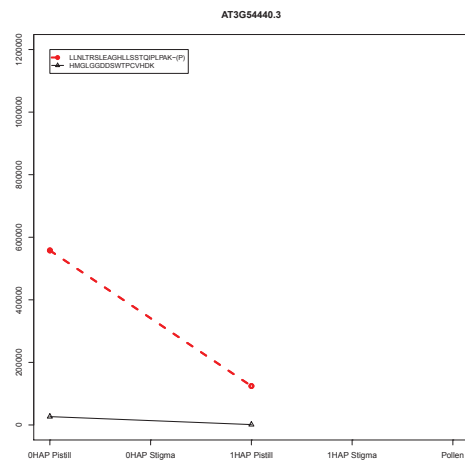
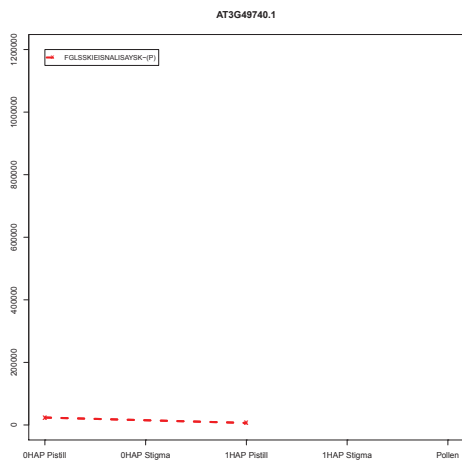
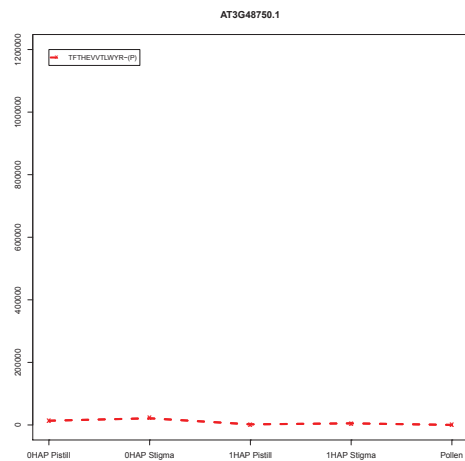
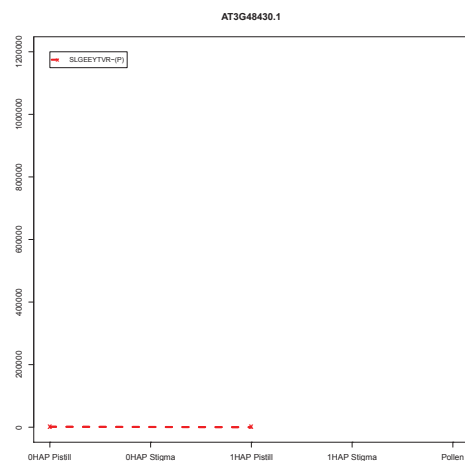
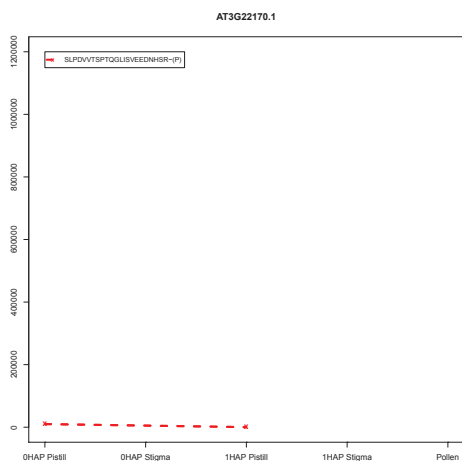
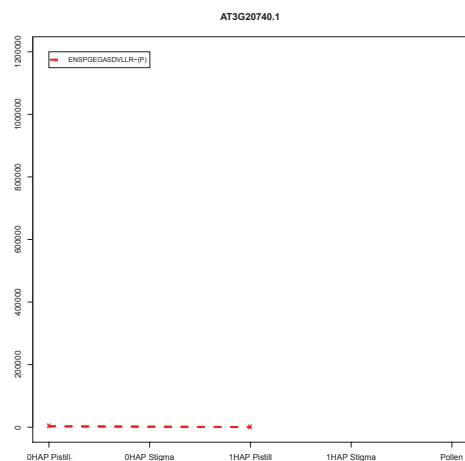




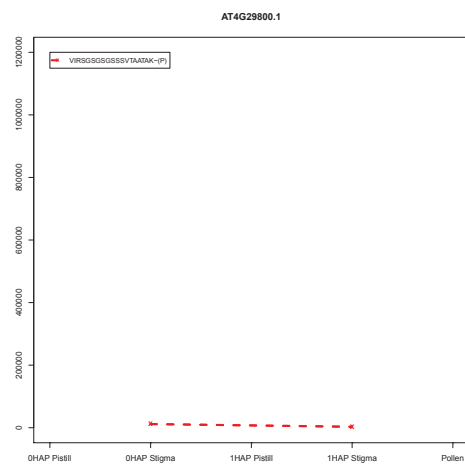
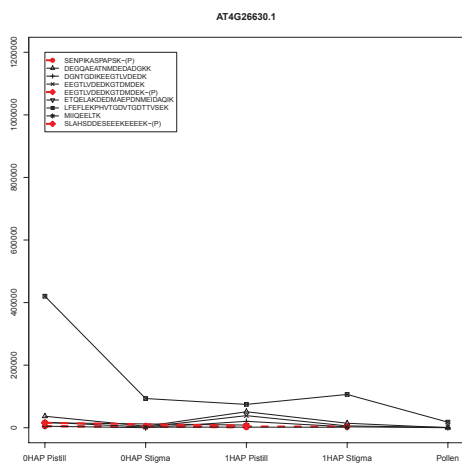
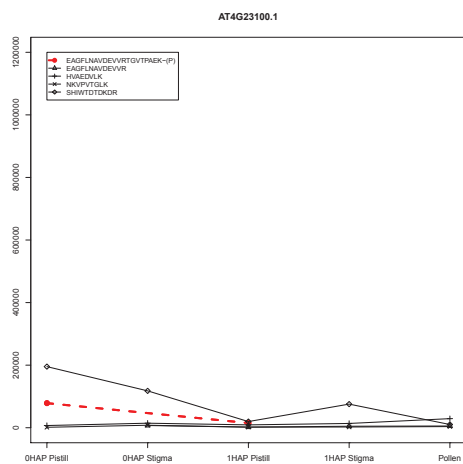
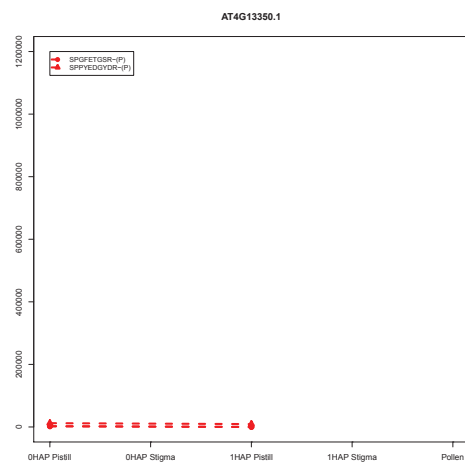
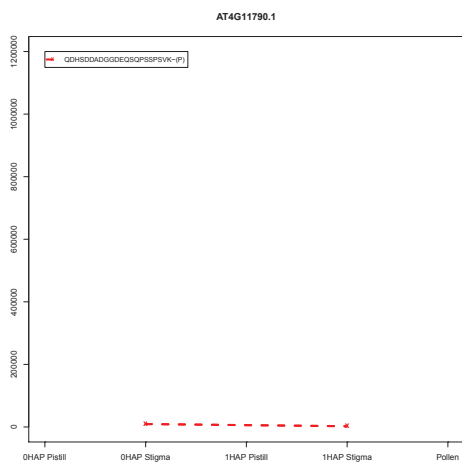
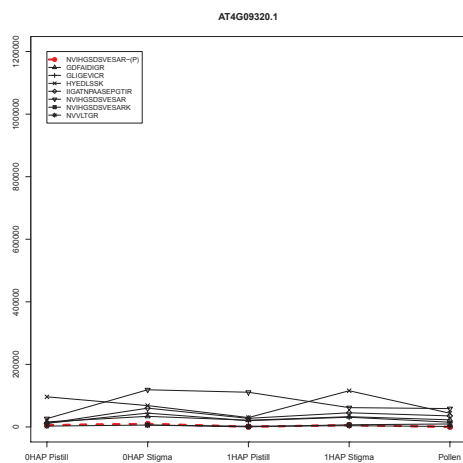
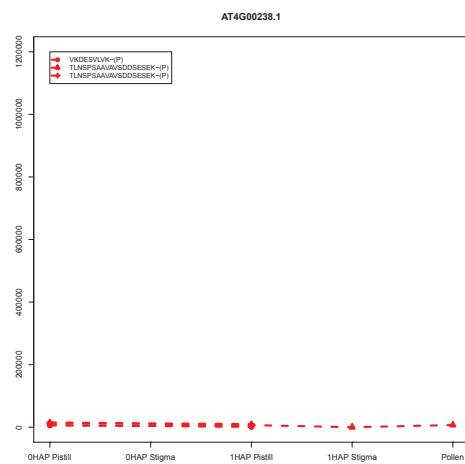
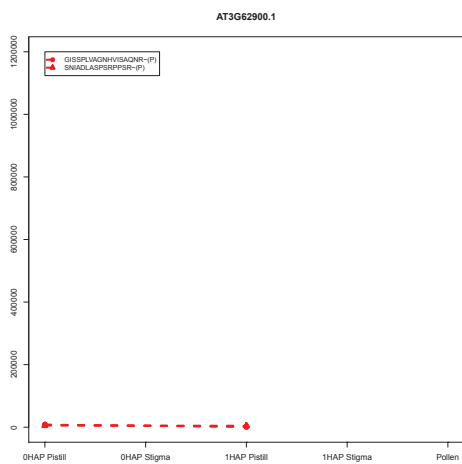
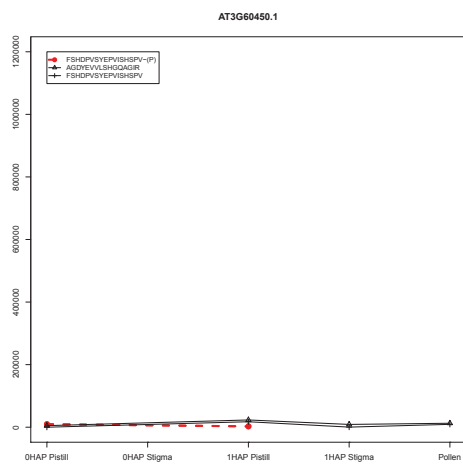


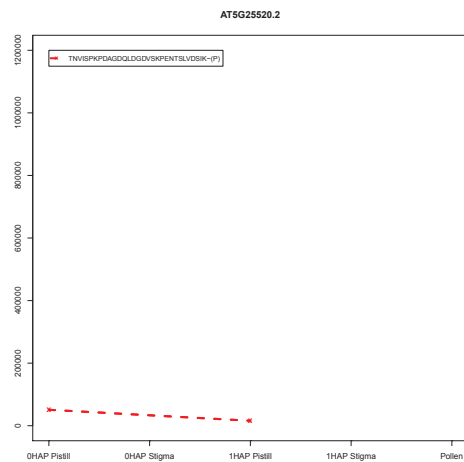
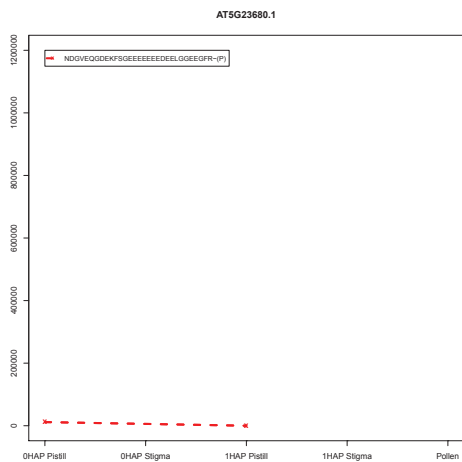
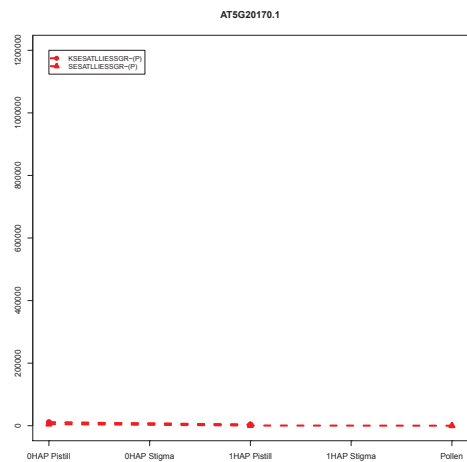
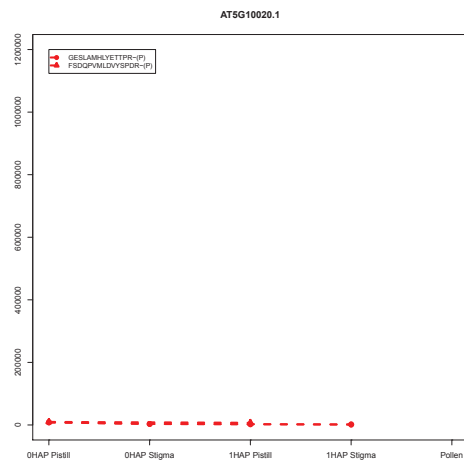
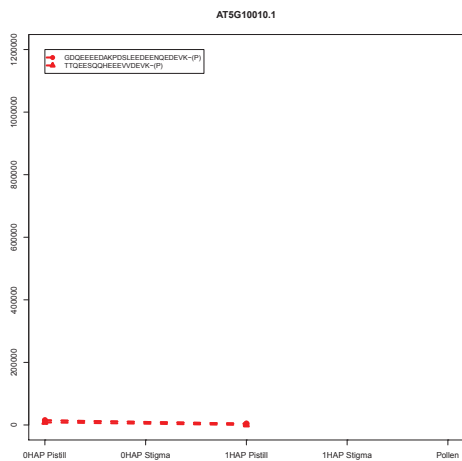
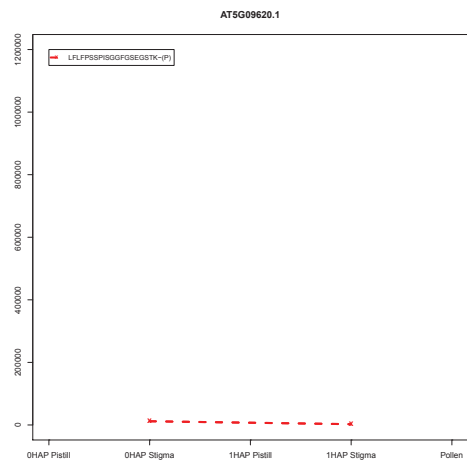
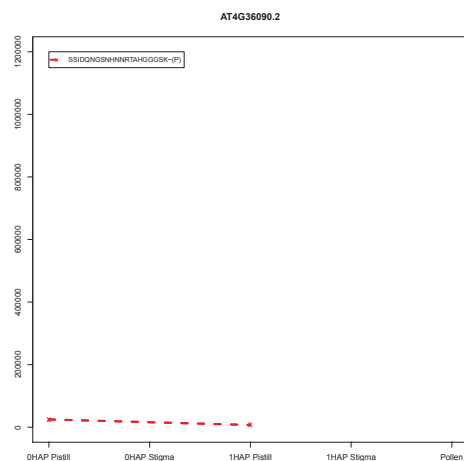
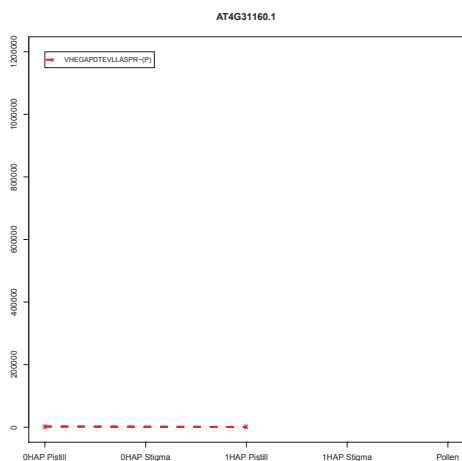
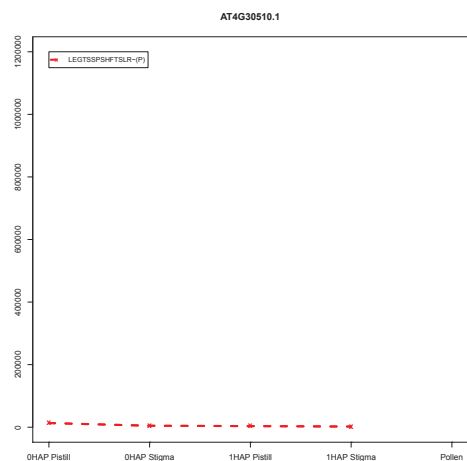


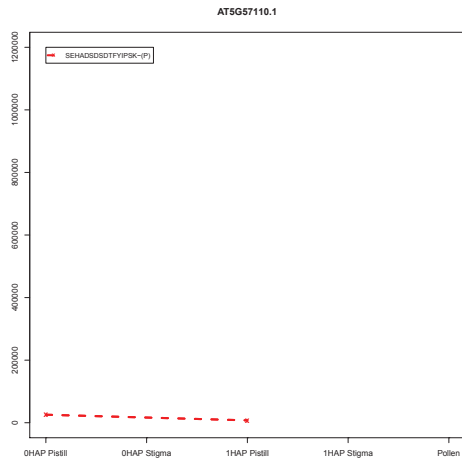
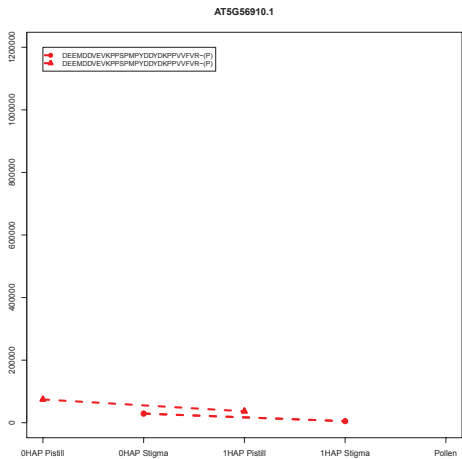
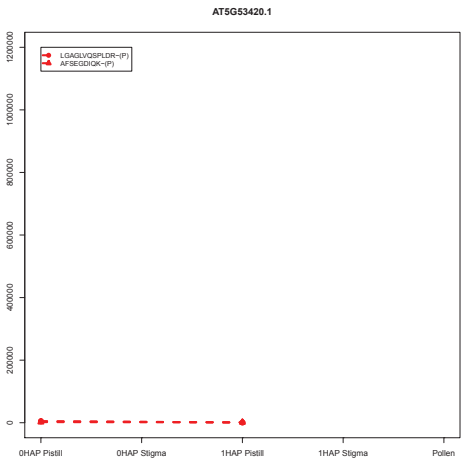
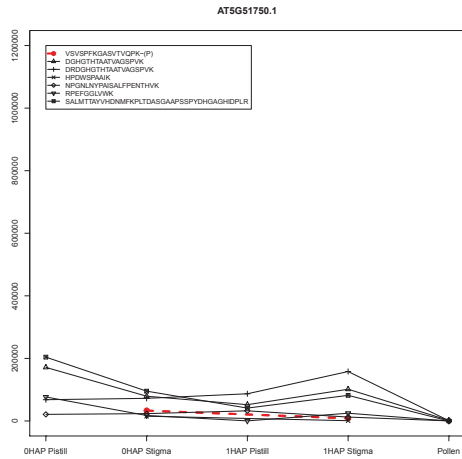
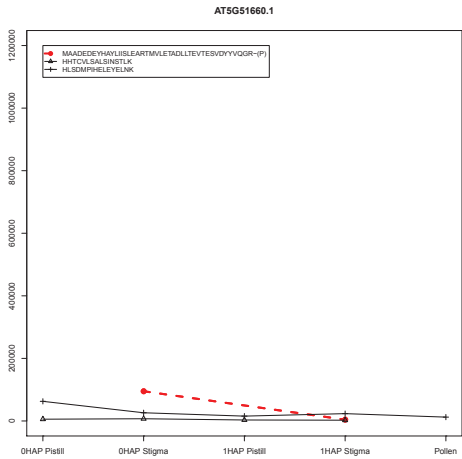
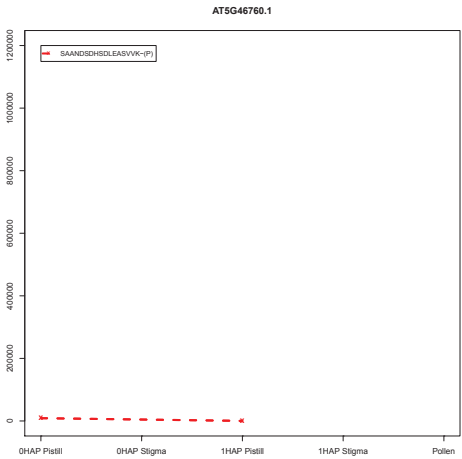
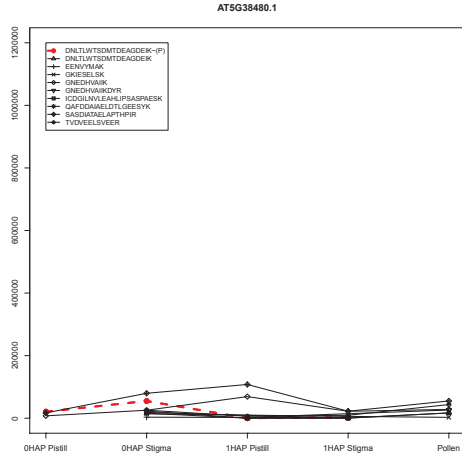
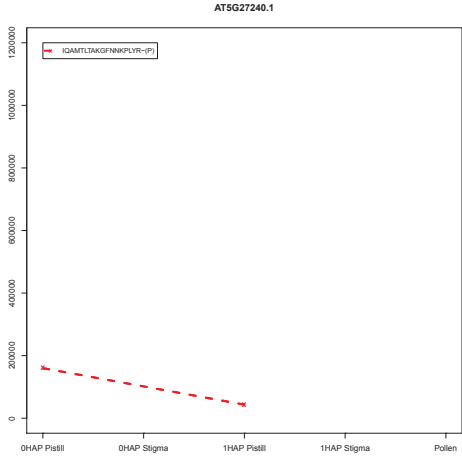
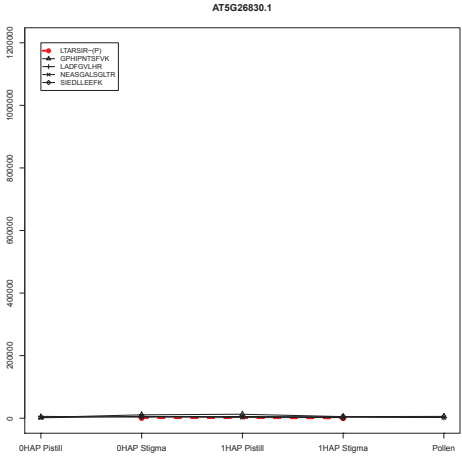


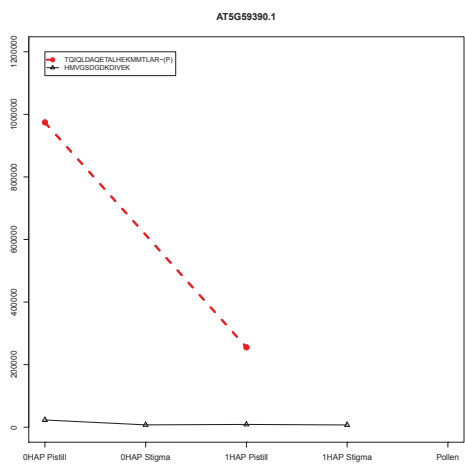












# Supplementary figure S1 d: Dephosphorylation candidate: male regulator of pollination

